(1)

AD-A207 512

U.S. Army Research Institute
for the Behavioral and Social Sciences

Research Report 1502

Impact of Excellence in Armor Program on Soldier Performance in One Station Unit Training

Raymond M. Mendel and Elizabeth S. Erffmeyer
Western Kentucky University

December 1988

89 5 03 074

BLANK PAGES
IN THIS
DOCUMENT
WERE NOT
FILMED

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract for the Department of the Army

Battelle

Technical review by

Barbara A. Black
Scott E. Graham

NOTICES

~~**DISTRIBUTION:** Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-POX, 5001 Eisenhower Ave., Alexandria, Virginia 22333-5600.~~

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

AD-A207 512

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS: --
2a. SECURITY CLASSIFICATION AUTHORITY: --	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE: --	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): --	5. MONITORING ORGANIZATION REPORT NUMBER(S): ARI Research Report 1502
6a. NAME OF PERFORMING ORGANIZATION: Battelle; 6b. OFFICE SYMBOL (If applicable):	7a. NAME OF MONITORING ORGANIZATION: U.S. Army Research Institute Fort Knox Field Unit
6c. ADDRESS (City, State, and ZIP Code): Research Triangle Park Office, 200 Park Dr., P.O. Box 12297, Research Triangle Park, NC 27709	7b. ADDRESS (City, State, and ZIP Code): ATTN: PERI-IK, Fort Knox, KY 40121-5620
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: U.S. Army Research Institute for the Behavioral and Social Sciences; 8b. OFFICE SYMBOL (If applicable): PERI-ZA	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: MIPR 22 ARI 88-86
8c. ADDRESS (City, State, and ZIP Code): 5001 Eisenhower Avenue, Alexandria, VA 22333-5600	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. 63743; PROJECT NO. A794; TASK NO. 3.3.1; WORK UNIT ACCESSION NO. C.6

11. TITLE (Include Security Classification)
Impact of Excellence in Armor Program on Soldier Performance in One Station Unit Training

12. PERSONAL AUTHOR(S)
Mendel, Raymond M., and Erffmeyer, Elizabeth S. (Western Kentucky University)

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 88-5 TO 88-9	1988, December	94

16. SUPPLEMENTARY NOTATION
Barbara A. Black, Contracting Officer's Representative.
Work was performed while authors were in residence.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Selection UCOFT
			Armor
			Training

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Promising Armor soldiers are enrolled in the Excellence in Armor training program (ET). This program is designed to accelerate a soldier's progression to tank commander by fostering early development of gunnery and related skills. The objectives of this research were to (1) determine if ET soldiers develop knowledge and skills beyond those of their normal track (NT) cohorts, particularly in the area of gunnery under both normal and degraded tank fire control system modes, (2) evaluate the degree of similarity between ET soldiers' and tank commanders' aptitude, interest, and temperament profiles, and (3) examine the validity of selected Armed Forces Vocational Aptitude Battery (ASVAB) and Project Alpha Predictor Battery (PAPB) scales for forecasting performance during one station unit training (OSUT).

Performance measures were developed to reflect the areas emphasized by the ET program of instruction (POI), the NT POI, and aspects common to both POIs. These measures were

(Continued)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION: Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL: Barbara A. Black	22b. TELEPHONE (Include Area Code): (502) 624-6928; 22c. OFFICE SYMBOL: PERI-IK

DD Form 1473, JUN 86 Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

ARI Research Report 1502

19. ABSTRACT (Continued)

administered to 83 ET soldiers and 83 NT soldiers matched on cognitive and psychomotor abilities. ASVAB and Project A predictor battery (PAPB) performance data were gathered for these soldiers and 41 Noncommissioned Officers (NCOs).

ET soldiers demonstrated performance gains over NT soldiers on measures targeting both the ET POI and the NT POI. On computerized armor training-device-based measures of gunnery performance, ET soldiers were more accurate and made fewer system management errors than did NT soldiers. These differences were traced to better performance on degraded exercises. Analyses of the relative similarity of ET, NT, and NCO ASVAB/PAPB profiles indicated that NT soldier profiles are more similar to NCO profiles than are ET profiles. One station unit training (OSUT) performance was predicted quite well by a combination of measures from the ASVAB and the PAPB.

Research Report 1502

Impact of Excellence in Armor Program on Soldier Performance in One Station Unit Training

Raymond M. Mendel and Elizabeth S. Erffmeyer
Western Kentucky University

Field Unit at Fort Knox, Kentucky
Donald F. Haggard, Chief

Training Research Laboratory
Jack H. Hiller, Director

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

December 1988

Army Project Number
2Q263743A794

Education and Training

FOREWORD

The U.S. Army Research Institute's Fort Knox Field Unit is committed to research that assists the Army in taking full advantage of new recruit capabilities. The Excellence in Armor program was initiated at Fort Knox in 1984. The purpose of the program is to identify high performing entry soldiers and to accelerate their training beyond the standard program of instruction. The program has become a model for other TRADOC Schools.

Results of this research provide objective data that support the existing program and recommend its expansion. Predictor scores from the Army's Project Alpha program can be used to enhance the Army's ability to identify candidate soldiers for this program before entry training.

This research was requested by the Armor School, was conducted under a Memorandum of Agreement titled "Continuation of the Training Technology Field Activity at Fort Knox, Kentucky," and was signed between Headquarters, Training and Doctrine Command (TRADOC), U.S. Army Armor School (USAARMS), and U.S. Army Research Institute (USARI) on 28 March 1987. The results have been briefed to the Assistant Commandant of the Armor School and have been used by the Office of the Chief of Armor to support the Deputy Chief of Staff's Quality of Accessions requirement.

EDGAR M. JOHNSON
Technical Director

IMPACT OF EXCELLENCE IN ARMOR PROGRAM ON SOLDIER PERFORMANCE IN ONE STATION UNIT TRAINING

EXECUTIVE SUMMARY

Requirement:

Promising Armor soldiers are enrolled in the Excellence in Armor training program (ET). This program is designed to accelerate a soldier's progression to tank commander by fostering early development of gunnery and related skills. The objectives of this research were to (1) determine if ET soldiers develop knowledge and skills beyond those of their normal track (NT) cohorts, particularly in the area of gunnery under both normal and degraded tank fire control system modes, (2) evaluate the degree of similarity between ET soldiers' and tank commanders' aptitude, interest, and temperament profiles, and (3) examine the validity of selected Armed Services Vocational Aptitude Battery (ASVAB) and Project Alpha Predictor Battery (PAPB) scales for forecasting performance during One Station Unit Training (OSUT).

Procedure:

Performance measures were developed to reflect the content areas emphasized by the ET program of instruction (POI), the NT POI, and the aspects common to both POIs. These measures were then administered to 83 ET soldiers and 83 NT soldiers matched on cognitive and psychomotor abilities. ASVAB and Project Alpha predictor batter (PAPB) performance data were gathered for these soldiers as well as 41 Noncommissioned Officers (NCOs).

Findings:

ET soldiers demonstrated performance gains over NT soldiers on measures targeting both the ET POI and the NT POI. On computerized armor training-device-based measures of gunnery performance, ET soldiers were more accurate and made fewer system management errors than did NT soldiers. These differences were traced to better performance on degraded exercises. Analyses of the relative similarity of ET, NT, and NCO ASVAB/PAPB profiles indicated that NT soldier profiles are more similar to NCO profiles than are ET profiles. One Station Unit Training performance was predicted quite well by a combination of measures from the ASVAB and the PAPB.

Utilization of Findings:

The data offer support for the continuation, perhaps even the expansion, of the ET program. Selection of ET soldiers and prediction of effectiveness during OSUT can be improved by using a combination of the ASVAB Combat Operations aptitude area selector score, the tracking factor score derived from the psychomotor component of the PAPB, and the combat scale derived from the PAPB.

IMPACT OF EXCELLENCE IN ARMOR PROGRAM ON SOLDIER PERFORMANCE IN ONE STATION UNIT TRAINING

CONTENTS

LIST OF TABLES

Page

LIST OF TABLES (Continued)

Page

LIST OF TABLES (Continued)

LIST OF FIGURES

APPENDIXES TO IMPACT OF EXCELLENCE IN ARMOR PROGRAM ON SOLDIER PERFORMANCE IN ONE STATION UNIT TRAINING*

CONTENTS

*Published under separate cover as a Research Note.

IMPACT OF EXCELLENCE IN ARMOR PROGRAM ON SOLDIER PERFORMANCE IN ONE STATION UNIT TRAINING

INTRODUCTION

In Fiscal Year 1984 the U.S. Army initiated the Excellence in Armor program, an accelerated training track at Fort Knox for M1 One Station Unit Training (OSUT) soldiers. This training, referred to as the Excellence Track (ET), identifies high performing OSUT trainees and accelerates their training program to include training beyond the standard Program Of Instruction (POI). Beginning with Week 8 of the 14 weeks of training, ET soldiers get more training on hard-skill tank tasks and technical subjects. To allow time for this additional training, the M1 OSUT Normal Training Track (NT) POI is presented to the ETs in compressed form. That is, ETs are expected to master the NT content, but must do so in less time. The time reclaimed in this manner is then devoted to additional training that is almost exclusively gunnery oriented. It therefore supplements the driving and loading training provided in the NT. Furthermore, ET soldiers serve as peer instructors for NT soldiers. This reinforces the training of the ET soldiers, although it makes it difficult to quantify the amount of training ET and NT soldiers receive.

The primary objective of this research effort is to evaluate the effectiveness of the M1 OSUT Excellence Track. In order to determine whether or not ET and NT soldiers differ in the skills acquired in OSUT, ET soldiers are compared to NT soldiers on a comprehensive set of performance measures derived from hands-on and paper-and-pencil tests as well as supervisor and peer ratings. These criterion measures were identified or developed to tap the ET domain, the NT domain, and the domain common to both programs. Thus, we were able to pinpoint precisely where ET/NT differences lie.

In order to maximize the effectiveness of an M1 tank, it is important to ensure that the best, most technically proficient tankers are selected to command the tank (Phillips, 1985). One purpose of the ET program is to accelerate the progression of high potential trainees. NT graduates typically come out of OSUT as an E-1 or E-2 (Loader). NT graduates typically require six to seven years to achieve Tank Commander (TC) (E-6). ET graduates generally come out of OSUT as an E-2 or E-3 and therefore are intended to progress to TC in only four or five years. In light of these progressions, it is of interest to determine if those trainees who are placed in the "fast track" to TC, that is, the ET graduates, have similar aptitude, interest, and temperament profiles to those of current TCs. This interest was predicated on the assumption that similar ET-TC profiles suggest that ETs will continue on to become TCs. Hence, the second major objective of this project is to evaluate the degree of similarity of the aptitude, interest, and temperament profiles of ET soldiers and NT soldiers to the profiles of senior Noncommissioned Officers (NCOs).

In addition to describing the differences between ET and NT soldiers' knowledge and skill acquisition, there is an interest in determining the variables that predict performance in OSUT. Thus, the third primary objective of this research is to identify performance parameters, both predictors and criteria, associated with high performance in initial entry training.

This report describes the methodology followed in accomplishing each of these objectives as well as the subsequent analyses and results. First, however, we review the literature relevant to our research objectives.

REVIEW OF THE LITERATURE

Past research addressing tank crew performance has focused almost exclusively on predicting loading, driving, or gunnery performance on the basis of cognitive, perceptual, psychomotor, or biodata measures. This past research is germane to our first research objective primarily as a source of information regarding criterion measures appropriate for evaluating ET and NT performance. A brief review of predictive studies is followed by a discussion of the criterion issues raised in these studies. With regard to our second research objective, the review of the literature helped to identify relevant dimensions on which to compare ETs and TCs. Thus, we conclude the review by exploring the literature linking aptitude, interest, and temperament measures to TC/Advanced NCO Course (ANCOC) performance.

Predictive Studies

Initial efforts to predict tank crew performance involved paper and pencil aptitude measures, primarily some sampling of Armed Services Vocational Aptitude Battery (ASVAB) subtests, as predictors. These early efforts were disappointing (Eaton, Bessemer, & Kristiansen, 1979; Greenstein & Hughes, 1977, cited in Campbell & Black, 1982). Eaton, et al. (1979) identified ASVAB and perceptual measures that related to OSUT performance for driving and gunnery. However, these relationships failed to cross-validate to soldiers in Table of Organization and Equipment (TOE) units. Eaton, et al. concluded that there was no support for paper and pencil tests as predictors of tank crew qualification gunnery. Likewise, Black and Mitchell (1985) surmised that paper and pencil tests have resulted in few significant relationships with gunnery performance for either trainee or TOE personnel. They suggested that paper and pencil tests tend to measure only cognitive or perceptual aptitudes and fail to assess the psychomotor aspects of gunnery performance. These discouraging findings gave impetus to the development of job sample tests as predictors of gunnery performance.

Job sample tests attempt to predict performance on the basis of actual samples of the behaviors that comprise job performance. Job sample tests have greater face validity and frequently result in significant relationships with performance criteria (Siegel & Bergman, 1975). Eaton, Johnson, and Black (1980) found that job sample predictors of gunnery performance validated on recent OSUT graduates failed to cross-validate to armor soldiers in TOE units. It was suggested, however, that these predictors might be useful as a basis for assignment to operational units after initial training. Biers and Sauer (1982) found that linear combinations of performance-based predictor measures across job samples accounted for a high proportion of the variability in Table VIII performance. Other experience-based and cognitive predictors evaluated by Biers and Sauer failed to show any relationships to the job sample predictors.

These studies suggest that job sample tests hold promise as predictors of tank crew performance. However, other studies have supported the validity of aptitude measures as predictors as well. Black (1980) found that the Combat Operations (CO) composite of the ASVAB, was related to TOE unit gunnery performance as reflected in Tank Crewman Readiness Tests for loaders and gunners. This relationship was not present for OSUT soldiers. Black concluded that CO may be a measure of trainability reflecting cognitive ability. Her findings suggested that during the elapsed time from the collection of OSUT measures to the collection of TOE measures, higher mental ability soldiers retained trained skills better than the lower mental ability soldiers.

Campbell and Black (1982) examined ASVAB subtests, biodata variables, and job sample tests as predictors of MI training success. The criterion measures included OSUT Gate II and Gate III Tests, instructor rankings, and Table VII gunnery performance. Regression analysis demonstrated that CO predicted training performance, i.e., Gate scores and rankings, better and more reliably than any other single predictor. Six job sample tests contributed to the CO prediction accuracy. The obtained validity coefficients, ranging from .33 to .76, were impressive. However, the authors concluded that until criterion measures can be adequately defined and more reliably measured, the predictive ability of job sample tests and biodata may be difficult to determine.

Black and Mitchell (1985) investigated the relationship among hands-on tests, computer-based tests, ASVAB subscores, motivation and experience. They found large differences in computer gunnery scores as a function of Armed Forces Qualification Test (AFQT) category. Experience as a gunner correlated with both hands-on tracking tests and hands-on target engagement tests. These predictor measures were correlated with supervisory ratings and Table VIII gunnery scores. None of the job sample tests correlated with the supervisory ratings. Likewise, there were no significant relationships between the computer-based predictors and the Table VIII measures. The

hands-on measures correlated with only two Table VIII night measures. Black and Mitchell attributed the failure to find relationships between the job sample predictors and the Table VIII measures to criterion problems.

Although there are some inconsistencies in these findings, generally they suggest that tank crew performance is related to certain aptitudes and abilities and can be predicted by appropriate measures. A problem common to many of these studies is a lack of relevance and reliability in the criterion measures. Another difficulty is the small sample size in a number of studies (Schmidt, Hunter, & Urry, 1976). Generally, the studies that had well developed criterion measures and adequate sample sizes were the studies that found significant relationships between the predictor measures and tank crew performance. The criterion problem argued forcefully for the development of more relevant, psychometrically sound measures of gunnery performance. This issue is discussed more fully in the following section.

Criterion Issues Identified in the Literature

The importance of sound criterion measures is clearly recognized in the literature (Black, 1980; Black & Mitchell, 1985; Campbell & Black, 1982; Eaton, et al., 1979, Graham, 1985). Criterion measures used in past studies include performance on live-fire gunnery tables, paper and pencil and GATE Tests from OSUT, Tank Crewman Readiness Tests, and instructor ratings. There are a number of concerns associated with the criteria used to evaluate tank crew performance, particularly those used to assess gunnery performance.

Eaton and Whalen (1980) documented the difficulty of obtaining accurately sensed live-fire measures. Under relatively good field conditions, the most accurate method (OSUT trainees with 10X periscopes and researchers with 7 x 50 binoculars) sensed only 87% and 86% of the rounds correctly. A frequently used scoring method, TCs using their M60A1 10X rangefinders, resulted in a very low 64% accuracy rate. This is even less impressive when one considers that 50% accuracy could be expected by chance.

Other sources of both unreliability and contamination in live-fire exercises have been identified. Variations in weather, tank equipment, range equipment, and ammunition characteristics inevitably result in increased error variance in the criterion measures (Graham, 1985).

It has been suggested that tank gunnery tables are not the most appropriate criterion measure for tests designed to predict combat criterion. Main gun live-fire, because of range safety constraints and the constraints of simulation, may not require the same type or same level of difficulty of tracking, round

sensing, target acquisition, and moving engagements that are a part of combat conditions. Thus, live-fire exercises are often deficient in this respect (Black & Mitchell, 1985).

An additional serious measurement concern with tank gunnery tables is that the measures provided are at the crew-level. It is not possible to determine the individual contribution of the driver, gunner, or TC to gunnery performance. Frequently ineffective crewmen are paired with experienced TCs to ensure that the tank crew will be rated as qualified while effective gunners are paired with poor TCs and fail to qualify their tank. For individual performance, therefore, the results of tank tables are likely inappropriate criteria (Black & Mitchell, 1985).

Performance ratings and hands-on criteria measures also have shortcomings. Hands-on performance tests are frequently constrained by time and equipment demands. The subjective nature of the performance rating process makes it particularly susceptible to bias. The process is complex and there are a number of influences that can affect the rating other than the performance of the ratee. These include rater characteristics, ratee characteristics, the rating instrument, organizational characteristics, and even the rating process itself (DeNisi, Cafferty, & Meglino, 1984).

Existing training evaluation methods such as paper and pencil tests and Gate Tests are also used as criterion measures. Morrison and Bessemer (1980) noted the importance of using appropriate criterion measures when evaluating training effectiveness. They found that some end-of-block (EOB) tests resulted in an excessive number of first-round NO-GOs largely because a written tests was being used to evaluate a performance skill. Even if the mode of testing is appropriate, existing training evaluation methods must be scrutinized for possible sources of contamination. Existing measures may be of limited value as trainees are sometimes familiar with the test questions prior to testing and are frequently coached for specific tests. This is particularly likely following a first round NO GO on a test. When such measures are used as criteria, care should be taken to eliminate these sources of contamination.

Black and Mitchell (1985) suggest that time and equipment constraints frequently result in researchers using criterion measures simply because they are readily available. They appeal for more emphasis on selecting measures of tank crew performance for their relevance and reliability.

Recent Developments in Criterion Measures

The Army recently developed a high fidelity computer-controlled simulator, the M1 Unit Conduct of Fire Trainer (UCOFT), that is designed to provide the necessary stimulus-response situations required to evaluate gunnery performance. Computerized simulators, such as the UCOFT, have the desirable

characteristics of precise presentation of target conditions with accurate scoring and timing. In addition, the UCOFT has the capability of presenting a variety of threat scenarios that systematically vary in degree of difficulty. The UCOFT also has the capability to test under degraded conditions. For these reasons, UCOFT performance scores were appealing criterion measures for evaluating gunnery performance across a range of skill levels.

Graham (1985) assessed the psychometric properties of various UCOFT-based gunnery scoring techniques. He found several measures, including Hit Rate and Target Identification (ID) Time, with stability coefficients above .80. Graham expressed two concerns regarding the UCOFT. One, it contains dispersion rounds which result in unreliability in performance measures. Secondly, there will likely be little variance in performance on easier engagements because of ceiling effects. Both of these defects were addressed through careful selection and scoring of UCOFT exercises in the present investigation.

In sum, the UCOFT has great potential not only as a measure of gunnery performance against which to validate predictor tests, but also as a basis for personnel placement decisions (Black & Mitchell, 1985). Accordingly, an additional objective of this research effort is to refine the UCOFT criterion measures developed by Graham (1985) and to use these measures as criteria in the evaluation of the ET program.

Literature Relevant to Assessing TC and ET Similarity

The second major objective of our project is to evaluate the degree of similarity of the aptitude, interest, and temperament profiles of ET soldiers to the profiles of ANCOC soldiers. The literature reviewed next influenced our selection of relevant aptitude, interests, and temperament dimensions on which to compare the soldiers.

The Gideon Report (Wallace, 1982) contains an assessment of the relationship between tank crewman mental ability (AFQT) and tank crew performance on live-fire gunnery. Wallace found a highly significant relationship between TC AFQT score and gunnery performance, but no parallel relationship between crew member AFQT and gunnery performance. Consequently, considerable attention has been focused on the relationship between TC mental ability and tank crew success. Black and Mitchell (1985) suggest that paper and pencil cognitive and/or perceptual tests are likely to be useful predictors of performance when the criterion task is more cognitively weighted as it was in the Gideon report. The live-fire task likely emphasized the where and when to fire (cognitive decision) rather than the psychomotor skills involved in how to fire. Consistent with this, Eaton, et al. (1979) found that their paper-and-pencil cognitive measures failed to cross-

validate to TOE TCs and gunners when a Table VIII criterion measure was used. Here prior knowledge of the Table VIII events effectively removed the cognitive demands of the exercises.

Job sample tests have also proven useful as predictors of TC performance. Eaton (1978, cited in Campbell & Black, 1982) found significant zero-order correlation coefficients between scores on a table-top tank gunnery simulator (Wiley Burst-on-Target Trainer), gunnery skills tests, and a mini-tank range and the criterion of armor tank crew (TC and gunner) score on the annual tank qualification exercise. Biers and Sauer (1982) report linear combinations of three computer-based and four hands-on job sample measures that explain a high proportion of the variability in past Table VIII performance of TCs and gunners.

Biodata variables have also been found to relate to gunnery scores of TCs. Successful TCs have been characterized as having more time in the TC position, more training time with their gunner, and a history of qualified tank crews (Black & Mitchell, 1985; Biers & Sauer, 1982).

In sum, the literature indicates that cognitive, performance, and non-cognitive measures, such as experience, are related to TC performance. Following from this, measures representing each of these domains were used in comparing ANCOC/TC-ET profiles.

EVALUATION OF THE EXCELLENCE TRAINING TRACK PROGRAM

This section describes our methodology for achieving the objectives of this project. This research effort centers around three primary objectives. The first objective is the evaluation of the ET Training Program. An important component of this evaluation is the comparison of ET-NT gunnery skill differences under conditions of degraded and non-degraded performance. The second objective is comparing senior noncommissioned officers with ET and NT crewmen. The third objective consists of delineating a model of performance parameters associated with high performance in initial entry training. Each of these objectives is addressed in subsequent sections of this report.

The following section deals with the first objective. The design considerations for ET Program evaluation are discussed first, followed by a discussion of the development and collection of criterion measures. The analysis of the criterion data and resulting findings are discussed.

Design Issues in the Comparison of ET and NT Soldiers

Rationale for the Design Selection

Selecting a reasonable research design for evaluating the effectiveness of the ET training program requires an understanding of the situational constraints in the research setting that might introduce bias into the measures collected or even restrict the type of data that can be collected. The rationale for our research design, Multivariate Analysis of Variance (MANOVA) with matching, is discussed below following a brief, general description of the concerns introduced by the ET selection process.

ET soldiers are selected from NT nominees recommended by the drill sergeants. Nomination for ET is based on a demonstrated ability to learn, motivation, military demeanor, and superior performance on certain tasks during the first seven weeks of OSUT. Non-random selection such as this typically reflects selector stereotypes and results in groups with reliable and substantial pre-existing differences. That is, if the same non-random selection process were repeated over and over again, the two groups would differ consistently in a number of ways. For example, since nomination for ET is based in part on OSUT performance, the two groups would likely differ on mean performance levels on certain OSUT tasks. It is quite probable that the two groups differ on a number of other variables, such as aptitude measures, as well. In short, ET and NT soldiers represent non-equivalent groups. This was of concern in designing the present research effort because these differences, quite apart from the training itself, may affect post-training scores. That is, selection differences may produce post-training differences between the groups even in the absence of a training effect. Therefore, to get a reasonable estimate of the impact of training, the analysis must properly control for these initial differences. That is, the effects of selection must be differentiated from the effects of training.

In MANOVA with matching using non-equivalent groups, our soldiers are matched on the basis of some pre-training measure(s) after the groups have been formed. Soldiers are paired so they have comparable scores on those measures. The matching process creates equivalent groups in the sense that variability about a common mean on measures judged to be relevant is equally distributed among the two training groups. The matching variables are not used as factors in the analysis (Cook & Campbell, 1979).

The rationale for using a matching design is straight forward: Since ET and NT groups are not equivalent, they cannot be directly compared. However, by including only those soldiers with similar scores on the matching variables, comparable groups are created and initial selection differences are controlled.

Identification of the Matching Variables

As indicated previously, the variables which differentiate ETs from NTs at the time ETs are selected are the most promising matching variables. However, a matching procedure is effective as a control for subject selection biases only if the matching variables are related to the criterion variables. ET and NT subjects were evaluated on a number of criterion measures. These measures were of four types: gunnery proficiency measures from the UCOFT, specific task measures from the Military Stakes and Tank Crew Gunnery Skills Test (TCGST), a paper and pencil test, and supervisory and peer ratings. The criterion measures are discussed in detail in the following section.

An important step in identifying and prioritizing matching variables was determining the current selection procedure for choosing ET soldiers. In addition to reviewing the official policy for ET selection (Phillips, 1985), subject matter experts (SMEs) were consulted to assist in identifying variables for the matching process. Sergeants were asked to describe the selection into the ET program. These descriptions were reviewed to determine how systematic the selection process is, that is, whether or not the same variables are used across companies. Although ET selection is based on scores from the ET Board Sheets, only the scores from specific Gate Tests and Basic Physical Fitness Test (BPFT) are systematic from company to company. The other selection variables identified by the sergeants included cognitive and psychomotor ability as well as intangibles such as motivation and leadership. Additionally, although ASVAB scores are not formally used as a selection variable, SMEs indicated they likely would be a relevant variable on which to match ET and NT subjects.

Cognitive ability and psychomotor ability, two variables deemed likely to correlate with the criterion measures, were selected as the matching variables. The Combat Operations (CO) composite from the ASVAB was determined to be the most relevant cognitive measure (Black, 1980; Campbell & Black, 1982). Factor 1, the most reliable scale score from the computer portion of the Project Alpha Trial Predictor Battery (PAPB) (J. J. McHenry, personal communication, March 1986), was used as the measure of psychomotor ability for matching purposes. Factor 1 is the mean of the log of the distance score for the two tracking tests contained in the PAPB. Factor 1 accounts for more variance than either of the other two psychomotor factors (J. J. McHenry, personal communication, March 1986). The split-half and test-retest reliabilities for the Tracking 1 Test (one-handed tracking) are .98 and .74, respectively, and for the Tracking 2 Test are .98 and .85, respectively.

Potential UCOFT Performance Contaminants

In addition to the problem of non-random selection discussed above, there are several other factors that required control through the design of the investigation. These included the potential effects of the confederates who served as TCs on the UCOFT criterion measures and potential systematic differences in the UCOFTs themselves.

Systematic effects of TCs were controlled by blocking. ET and NT soldiers were randomly assigned to each TC such that each confederate served as TC for an equal number of subjects from each group, thus avoiding treatment comparison bias since there was equal representation of the TC source of variability in each group. In addition, blocking was entered into the analysis such that this source of systematic variation was removed from residual error, providing an unconfounded test for training and TC effects. It should also be noted that although each TC tested an equal number of ET and NT subjects, the TC did not know which soldiers were ETs.

Any systematic differences between UCOFTs used in this investigation were controlled by counterbalancing the assignment of TC to UCOFT in such a manner that each TC spent equivalent time on each of the UCOFTs. This balanced any UCOFT effects across TCs so that they did not differentially influence performance of the soldiers. As in the case of our TCs, UCOFT operators did not know which subjects were ETs and NTs. The assignment of ET and NT subjects was also counterbalanced between the UCOFTs. That is, ET and NT subjects were randomly assigned to each UCOFT in equal number. Thus, there should be no UCOFT induced differences between the training groups.

In sum, MANOVA with matching on cognitive and psychomotor ability was the technique used for comparing the effectiveness of the ET and NT Training Programs. Matched ET soldiers and NT soldiers are compared on criterion measures from the UCOFT, Military Stakes, TCGST, Paper and Pencil Test, and Project Alpha (Project A) Ratings. Measures from four of the five performance domains were separately analyzed using a one-way MANOVA procedure. The UCOFT measures were analyzed by a two-way MANOVA with blocking on TC. Significant MANOVAs were followed by univariate analyses of variance (ANOVAs) to determine the source of the variation in the criterion measures. These analyses are described in detail in the Analyses and Results Section of this report. A detailed description of the criterion measure development follows a brief discussion of the power analysis used to determine the optimal sample size for this evaluation. This is then followed by the more detailed description of the data analysis.

Power Analysis

Sample size requirements needed to achieve various levels of statistical power have been developed for various training intervention designs broken down by effect size and alpha levels (Arvey, Cole, Hazucha, & Hartano, 1985; Asher & Sciarrino, 1981). However, before a investigation is conducted, the researcher often has little basis for estimating the expected effect size. Since effect size is the major determinant of the sample size requirement for a given power level, this missing information is often a serious problem.

Asher and Sciarrino have provided a helpful solution to this problem. They surveyed more than 200 training studies published between 1960 and 1981, tabulating the magnitude of the reported training effect. The twenty-five percent of the studies reporting the largest training effects were classified as "large" and their median effect size reported. This was repeated for the middle 50% and the bottom 25% of the studies. In this way Asher and Sciarrino have provided a historical basis for estimating training effect size. Assuming a "medium" effect size, a power of .80, and an alpha level of .05, 37 subjects are required in each training group. This number is greater than the 19 subjects required to detect a "large" effect, but smaller than the 380 required for detecting a "small" training effect.

As a compromise between the small and medium effect size sample requirements, we sampled 166 soldiers, 83 ETs and 83 NTs. Our research participants are described below.

Research Participants

All research participants in the NT-ET comparison were US Army enlisted personnel undergoing basic Armor training in the 1st Armor Training Brigade, 1st Battalion, Armor, Fort Knox, Kentucky. One hundred and sixty-six soldiers (Military Occupational Specialty (MOS) 19K) were drawn from A, B, and C companies across five cycles between May and December 1986. All five cycles were subsequent to the ET POI implemented in May 1986.

A two-step procedure was followed to select participants from each cycle of OSUT. Immediately following ET selection in Week 8 of OSUT, all ET soldiers and, for each ET, two NT soldiers matched on the Combat Operations (CO) Composite of the ASVAB were administered the Project A Trial Predictor Battery (PAPB). The number of NT soldiers was subsequently reduced by one-half through matching each ET soldier with a single NT soldier based on CO and the Factor 1 score from the PAPB psychomotor tests. Thus, the final research group from each cycle consisted of 83 ET soldiers and an equal number of NT soldiers matched on cognitive (CO scores) and psychomotor (Factor 1 scores) abilities.

As shown in Table 1, ET and NT participants had nearly identical mean values on both of the matching variables. Thus, the matching process was successful in creating equal groups on these critical variables as well as on the other ASVAB composites and Project A Psychomotor measures (see Table 1).

Table 1. Mean ASVAB and Psychomotor Scores for ETs and NT Matches

ASVAB COMPOSITE	ET	NT
GT		
Mean	111.34	111.82
Std. Dev.	10.54	9.50
GM		
Mean	113.76	113.08
Std. Dev.	13.91	11.98
EL		
Mean	112.63	112.25
Std. Dev.	13.85	11.93
CL		
Mean	111.04	110.90
Std. Dev.	12.81	10.80
MM		
Mean	115.24	114.88
Std. Dev.	11.87	9.95
SC		
Mean	114.36	114.33
Std. Dev.	11.34	9.93
CO		
Mean	115.46	115.48
Std. Dev.	10.77	9.62
FA		
Mean	113.34	112.93
Std. Dev.	12.81	10.24
OF		
Mean	114.68	114.26
Std. Dev.	10.45	8.56
ST		
Mean	113.10	111.87
Std. Dev.	13.70	11.42

Table 1. Mean ASVAB and Psychomotor Scores for ETs and NT Matches (Cont.)

PROJECT A PSYCHOMOTOR TESTS	ET	NT
FACTOR 1		
Mean	-.724	-.714
Std. Dev.	.752	.696
TRACKING 1		
Mean	-.689	-.693
Std. Dev.	.693	.634
TRACKING 2		
Mean	-.758	-.735
Std. Dev.	.940	.832
FACTOR 2		
Mean	-.388	-.419
Std. Dev.	.677	.497
CANNON		
Mean	-.395	-.357
Std. Dev.	.799	.780
TARGET 2		
Mean	-.384	-.485
Std. Dev.	.846	.574
FACTOR 3		
Mean	-.294	-.238
Std. Dev.	.928	.889

Criterion Measures Overview

Based on reviews of the content of both the ET and NT Training Tracks, criterion measures were identified and/or developed to reflect training program content (1) to which NTs are more exposed, (2) to which ETs are more exposed, and (3) to which ETs and NTs have comparable exposure. A number of performance measures and written tests were used as criteria to document what is learned in the training tracks. These measures are introduced very briefly immediately following. Each measure is then described in more detail in the subsequent section.

NT training content was reflected by performance on the Military Stakes and on a paper and pencil knowledge test. The Military Stakes measure included total time for the Military Stakes course, three Army-administered stations with adequate variability to differentiate between soldiers, and two stations tested via simulation by project staff.

ET training content was reflected by performance on the TCGST and on gunnery exercises on the UCOFT. To obtain TCGST measures, NT soldiers were tested by 1st Battalion concurrently with the standard ET testing on 13 of the 18 TCGST stations. Three additional stations were tested via simulation by project personnel. The UCOFT measures consisted of a 2-hour session during which six exercises were administered twice. The exercises selected were based on the input of 1st Battalion staff to represent a range of difficulty. Soldiers participating in our investigation did not have any exposure to the UCOFT prior to our testing session.

Performance rating scales developed for Project A were used to reflect common ET and NT training content. All participants were rated by peers and drill sergeants on the Project A Army-Wide Rating Scales.

Schedule of Criterion Data Collection

The TCGST is normally administered EOB as part of the ET Program during Weeks 10 - 13 of OSUT. Likewise, the Military Stakes are administered as a matter of course in Week 13 of OSUT. These tests were conducted following the standard 1st Battalion schedule. During Week 13 all ETs and NT matches participated in a four-hour testing session conducted by ARI project personnel. During this session the paper and pencil knowledge test, the Project A Peer Ratings, and the TCGST and Military Stakes simulations were administered. The UCOFT exercises were administered in individual testing sessions during Weeks 13 and 14. The Project A Supervisory Ratings were collected during a session held for NCOs during OSUT Week 13 or 14.

All criterion measures were pilot tested on two cycles of ET/NT soldiers prior to the data collection on the five cycles reported herein. The pilot administration of the measures and the data collected provided information useful for the refinement of the various criterion measures and the testing process.

Criterion Measures: NT Domain

Military Stakes

NT soldiers routinely take two final hands-on tests, Gate III during Week 10 and the Military Stakes during Week 13. Graduation from OSUT is conditional upon passing these tests. Gate III is a test of the tank skills required for effective operation of the driver, loader, and gunner stations. The Military Stakes is a test of non-tank subjects which is conducted at substations along a 5-mile course that must be completed within a specified time period. These tests are administered by cadre not affiliated with the companies being tested. As such, test administrators are not accountable for a particular

soldier's performance nor are they aware which soldiers are part of the ET program. The timing of these tests, the importance attached to performance thereon, and the relatively neutral conditions of test administration argued for their inclusion as criterion measures.

Concerns Regarding Military Stakes and Gate III

For our purposes, Gate III and Military Stakes performance measures had some shortcomings. These included a ceiling effect resulting in substantially restricted score variance, masked between soldier variation in training time required to prepare a soldier for the tests, variations in test content from one company to another, and differential sampling of the ET versus the NT content domain. These issues and the steps taken to address these shortcomings are discussed below.

Restricted Score Variance. Perhaps the most troublesome problem was the restricted variance in Military Stakes and Gate III scores. Each task is scored pass/fail (i.e., GO/NO GO). Soldiers are offered three opportunities to pass each task. Literally 100% of the soldiers in our two pilot companies received GOs on every Gate III task. Approximately 90% passed each task on the first administration. We proposed increasing score variance by modifying the scoring procedure for both of these tests. However, the Battalion was not receptive to any modification to the Military Stakes scoring procedure. This meant the Military Stakes data consisted of the GO/NO GO score on each task and the total time to complete the course.

Gate III Modifications. The Battalion agreed to two supplements to the Gate III scoring procedure that were intended to increase score variability. Morrison and Bessemer (1981) found that simply including execution times on tank tasks revealed differences not reflected in the dichotomous GO/NO GO scores. Thus, the scoring procedure for each timed task was modified to include recording the exact time of completion.

In addition, we observed Gate III administration and met with test administrators and cadre to identify a relevant dimension on which task performance could be evaluated equally well across all stations. The SMEs defined proficiency in terms of familiarity with the training manual. Gate III tasks are delineated step by step in the training manual, which soldiers are permitted to review during test administration. Soldiers who know the training manual well are able to complete the tasks more quickly and with greater facility. Thus, a 5-point rating scale was developed to measure familiarity with the training manual. Behavioral anchors were developed for the two extreme and the middle ratings. The Gate III score sheets were modified to include a place for recording the exact time for timed tasks and for rating familiarity with the training manual for each station. The modified score sheets may be found in Appendix A (published in separate Research Note). Test administrators attended a training session in which they were instructed on how to use the

rating scale and the exact time measures. Refresher training was provided prior to the administration of Gate III for each cycle. In addition, the Battalion agreed to use the same version of the Gate III test for the duration of our data collection.

<u>Discontinued Use of Gate III</u>. Subsequent monitoring of the Gate III data collection revealed that irregularities in the administration of the test rendered these data meaningless for our purposes. Thus collection of Gate III scores was terminated after the third cycle and the data were not analyzed.

<u>Masked Between Soldier Variation in Training Time</u>. A deficiency of the Military Stakes is its failure to reflect the amount of training resources and effort that are devoted to bringing a soldier up to passing performance level. Soldiers who perform inadequately typically receive considerable additional training prior to the test administration. Moreover, a distinguishing feature of ET training is "compression", that is, the practice of providing ETs the same training NTs receive in considerably less time, thereby creating the time necessary for additional training not possible with the NTs. Thus, although the Military Stakes is designed to tap the basic skills required of all 19Ks, a considerably smaller portion of ET training time is devoted to many Military Stakes subject areas. In short, equivalent Military Stakes scores often do not reflect comparable antecedent behaviors. ET performance equivalent to NTs on non-tank Military Stakes tasks reflects, in one sense, performance superior to NTs since proportionately fewer resources are allocated to their mastery of the component tasks.

<u>Differential Sampling of ET/NT Training Domains</u>. We attempted to partition the Military Stakes tasks to reflect differences in the content of ET and NT training to provide a more sensitive and relevant index of behavioral consequences of the two training programs. Our intention was to create three composites, each representing either NT training, ET training, or training common to both ET and NT programs. Five cadre SMEs, thoroughly familiar with the content of both the NT and ET training POIs, sorted the Military Stakes tasks into one of three categories: (1) tasks on which NTs receive more training (NT composite), (2) tasks on which ETs receive more training (ET composite), or (3) tasks on which ETs and NTs receive equivalent training (common composite). The tasks contained in each category form the three composites. However, examination of the Military Stakes data revealed that only three of the nineteen tasks had adequate score variance to warrant further analyses. Thus, development of any POI specific composites for Military Stakes measures was not possible.

<u>Military Stakes Measures</u>

Only three of the sixteen Military Stakes tasks resulted in ten or more first-round NO GOs across all five cycles of ET and NT soldiers. These tasks are Station 1: Estimate Range, Station

8: Perform Operator's Maintenance on a Caliber .45 Pistol, and Station 10B: Perform Operator's Maintenance on the M16A1 Rifle. The other Military Stakes tasks lacked adequate variability and were dropped from further analyses.

In addition to the cadre administered Military Stakes, two stations were administered via paper and pencil simulations by Army Research Institute (ARI) project personnel during our Week 13 testing session. These were Station 4: Recognize and Identify Friendly and Threat Armored Vehicles and Station 6: Visually Identify Potential Threat Aircraft. These simulations were developed by having cadre SMEs select slides of vehicles (Armored Vehicle Recognition, 1984) and aircraft (Aviator's Recognition Manual, 1977; Visual Aircraft Recognition, 1983) similar to those that appear on the Military Stakes test. In the Station 4 simulation, soldiers were required to indicate whether each of 20 vehicles was friendly or threat. In the Station 6 simulation, soldiers were required to record the numerical designation or standard NATO reporting name for each of eight aircraft. The score for each station was the number of items answered correctly. The two simulations may be found in Appendices B and C (published in separate Research Note), respectively.

In sum, the resulting measures for the Military Stakes were first-round GO/NO GO scores for three cadre administered tasks, the total time to complete the 5-mile Military Stakes course, and the two scores from the Station 4 and Station 6 simulations. The time measure was reflected so that a higher score indicates better performance. Each of the six Military Stakes measures was converted to a standardized T score (i.e., $M = 50$; $SD = 10$) prior to data analysis.

NT Paper and Pencil Test

A paper and pencil knowledge test was developed based on the NT POI. The test content represents Weeks 8 - 14 of the NT training content, that is, the training period concurrent with the ET program. A written test was included as a criterion measure for several reasons: the Army-administered performance measures allow trainees to refer to manuals and therefore do not measure knowledge retention; some of the classroom training content is never practiced in a hands-on setting, making performance tests less appropriate; and a written test is more economical than performance tests for assessing knowledge of a wide range of material. The test development effort is described below. A more detailed account of this effort may be found in Seibert (1987). Seven cadre SMEs, actively involved with OSUT training, contributed significantly to the test development.

NT Test Development

The content of the training POI for Weeks 8 - 14 was determined by examining training schedules, lesson plans, and training manuals. Thirty-nine distinct lessons or tasks were identified. Five of these are field exercises in which material learned previously is practiced. Since the content of these five exercises is tested in Gate tests and does not readily lend itself to paper and pencil testing, these five exercises are not included in the NT Paper and Pencil Test (NTPP).

Each SME independently estimated the amount of training time NT soldiers receive on each of the 34 lessons. The SMEs also rated the relative importance of each topic on a five-point graphic rating scale ranging from "not very important" to "extremely important". Interrater reliability was .83 for the importance rating and .64 for the time estimates. The lower reliability of the time estimates may have been due to some confusion among SMEs as to whether time included "doing" and "observing" or just "doing" a task. These data guided the inclusion of items to ensure the appropriate proportional representation of topics on the test.

The SMEs, following a training session on item writing, wrote two or three test items for up to ten topics. A given topic was assigned to one, two, or three SMEs based on its mean importance ratings. Due to the low interrater reliability and apparent confusion about the meaning of the time estimates, they were not used in determining test content. In addition, EOB tests, used for experienced soldiers as they are retrained from other MOS for M1 crew member duty, were reviewed to search for appropriate items. A pool of 267 preliminary items was generated by these efforts, 210 items written by SMEs and 57 items culled from existing tests.

The SMEs assisted in the review of these preliminary items by editing and clarifying items; by verifying that there was one and only one correct answer for each item or modifying the response options until this was so; by independently estimating the percent of trainees that would pass each item; and by selecting the better item of duplicate items. Items were eliminated by three criteria, that is, a lack of unanimous agreement on a single correct answer, the item duplicated the content tested by another item, or the estimated mean pass rate for the item was greater than 80% or less than 20%. This process narrowed the item pool to 197 items.

The remaining items were divided into two pilot tests for pretesting. Each pilot test was administered in a two-hour session to a group of 25 or 27 soldiers in their 14th week of OSUT. Item-total correlations and item difficulty indices (i.e., percent of examinees passing the item) were calculated for each of the 197 items. Item-total correlations ranged from -.36 to .75 with a mean of .18. Item difficulty ranged from .07 to 1.00 with a mean of .49. An item was retained for the final

version of the test if its item-total correlation exceeded .25 and its difficulty index was between .20 and .80. These guidelines were violated to ensure appropriate representation of all training areas by including nine items with item-total correlations as low as .17 and difficulty indices ranging from version of the test, with only two of the 34 topic areas under represented. The NT Paper and Pencil Knowledge Test appears in Appendix D (published in separate Research Note).

Administration of the NTPP

The NTPP was administered during the four-hour testing session held for all NT and ET participants during Week 13. Soldiers had one hour in which to complete the test.

NTPP Reliability, Item Statistics, and Scoring

The internal consistency of the NTPP, as measured by Cronbach's alpha, is .79. Item difficulty ranged from .16 to .92 with an average of .57. Item-total correlations ranged from -.04 to .42 with a mean of .20. Two items had negative item-total correlations (Item 19 $r_{it} = -.02$; Item 60 $r_{it} = -.04$).

The score on the NTPP is the number of items answered correctly. The mean number of items answered correctly across all five OSUT cycles is 42.95 (SD = 8.29). For purposes of data analysis, the NTPP scores are converted to standardized T scores.

Criterion Measures: ET Domain

UCOFT-Based Measures

A series of UCOFT exercises were presented to the ET and NT soldiers during their thirteenth or fourteenth week of OSUT. The purpose of this task was to gather measures of each soldier's gunnery skills under the uniquely standardized conditions afforded by the UCOFT. Accordingly, with a trained confederate serving as TC, each soldier from the gunner's station attempted to "destroy" a number of computer-generated targets.

Selection of the UCOFT Exercises

Presently, more than 700 UCOFT exercises are available. The more than 300 exercises designed for the simultaneous training of TCs and gunners form a matrix of combat conditions potentially useful as exercises for gunnery skill evaluation. The dimensions along which combat conditions can be manipulated through choice of exercises include the sight visibility, target range and number, systems malfunctions, distractions, and own vehicle and target movement.

This matrix was reduced based on information drawn from Graham's (1985) initial investigation on the psychometric properties of the UCOFT and input from a panel of six cadre SMEs familiar with the UCOFT and the NT and ET training tracks. Exercises that were judged to be too difficult (i.e., 80% or more of ETs and NTs would be expected to fail the exercise) and exercises that were judged to be too easy (i.e., 80% of more of the ETs and NTs would be expected to pass the exercise) were eliminated, leaving 111 exercises under consideration for inclusion.

SMEs were also asked to indicate for each UCOFT exercise whether, due to the content of the training POIs, ETs would perform better than NTs, NTs would perform better than ETs, or ETs and NTs would perform equally well. From the remaining 111 exercises, six were selected to form a representative cross-section of the gunnery skills taught in the ET POI. The six selected exercises (Exercises Number 322230, 311610, 322420, 325120, 313510, 314520) included four for which SMEs indicated ETs should out perform NTs and two for which SMEs indicated ETs and NTs should perform equally well. The selected exercises were also representative of the dimensions contained in the UCOFT exercise matrix, including 79% of the combat conditions in the complete UCOFT matrix. The engagement conditions represented in the final UCOFT gunnery skills test are shown in Table 2.

Table 2. Engagement Conditions in the UCOFT Gunnery Skills Test

U-COFT EXERCISE	PURPOSE	OWN VEHICLE	NUMBER	TARGET KIND	TARGET RANGE	VISIB	ENGAGE MODE	OPTICS	FIRE CONT HALF	FIRE CONT MODE
111210	Practice	Stat	Single	Stat	1500m	Day	Prec	GPS/Day	None	N
313110	Practice	Stat	Single	Moving	1500m	Day	Prec	GPS/Day	None	N
322230	GST	Stat	Single	Stat	1500m	Night	Prec	GPS/TIS	None	N
313510	GST	Stat	Single	Moving	1500m	Day	B.S.	GPS/Day	LRF COAX	N
311610	GST	Stat	Single	Stat	1500m	Day	B.S.	GAS/Day	LRF COAX GPS COMP	C
314520	GST	Moving	Single	Stat	1500m	Day Fog	B.S.	GPS/TIS	LRF COAX	N
322420	GST	Stat	Single	Stat	1500m	Night	Prec	GPS/TIS	STAB COAX	E
325120	GST	Moving	Single	Moving	1500m	Dusk	Prec	GPS	None	N

Note: Stationary (Stat)
Gunnery Skills Test (GST)
Gunner's Primary Sight (GPS)
Normal (N)
Thermal Imaging System (TIS)
Precision (Prec)
Battlesights (B.S.)
Emergency (E)
Laser Rangefinder (LRF)
Stabilization (STAB)
Computer (COMP)
Gunner's Auxiliary Sight (GAS)

Graham (1985) had discovered that a ceiling effect on certain UCOFT measures resulted in restricted variance on these measures. Pilot testing of our exercises with ten recent M1-OSUT graduates indicated adequate variance on the six UCOFT measures Graham determined to be sufficiently reliable to evaluate gunnery performance, i.e., Hit Rate, Target ID Time, Opening Time, Target Acquisition Composite, and Reticle Aim Composite.

TC Confederates

Two retired NCOs and a project staff member served as TC confederates. The TCs received 18.5 hours of training in the UCOFT as tank commander. They received an additional 14.5 hours of training in the UCOFT as gunner. TCs, as well as the UCOFT operators, were instructed not to correct mistakes or advise participants during test sessions. TC performance was monitored through the use of audio tapes. By the end of training, each TC had better than 90% accuracy in their fire commands for the selected engagements. That is, over 90% of the fire commands were flawless across both presentations of the exercises. Prior to the UCOFT testing for each cycle of soldiers, TCs received four to six hours of refresher training. Continued monitoring of TC performance ensured that at least a 90% level of accuracy was maintained throughout the data collection.

Test Sessions

The UCOFT exercises were administered individually in two-hour sessions during Week 13 or 14. The test sessions consisted of a brief review of the gunner's controls, the presentation of a target familiarization scenario, the presentation of a practice scenario (the first five engagements of Exercise 313110), the presentation of the first five engagements from each of the six selected exercises, a brief rest break, and a second presentation of the first five engagements of the six selected exercises. Thus, the six exercises that comprised the UCOFT measure were administered twice to each participant during the test session.

Dependent Measures

Eight performance measures were obtained from each UCOFT engagement. These are Hit Rate, Azimuth and Elevation Errors, Target Identification (ID) Time, Opening Time, and the three UCOFT composite measures (i.e., Target Acquisition, Reticle Aim, and System Management). These measures are described briefly below and in greater detail in the Unit-Conduct of Fire Trainer Instructor's Utilization Handbook (1985).

Hit Rate is a measure of whether or not the round hit the target. Azimuth Error is total distance in mils of round left or right of target center mass. Elevation Error is the total distance in mils of round above or below target center mass. Target ID Time is the time in seconds from the appearance of the target until the gunner identifies it. Opening Time is the time in seconds from the appearance of the target until the gunner fires the first round.

Target Acquisition is a composite of Target ID Time and Identification and Classification errors (i.e., the number of times during each exercise the gunner fails to identify or falsely identifies a target). Reticle Aim is a composite of Opening Time, Azimuth Error, Elevation Error, and Time to Kill (i.e., the time in seconds from the appearance of the target until the gunner hits the target). System Management is a composite of pre-firing switch errors, ammunition selection errors, and excessive own vehicle exposure time. Target Acquisition and Reticle Aim are reported as a letter grade of A, B, C, D, or F, with corresponding numerical values of 4.0, 3.0, 2.0, and 1.0. System Management is reported as a letter grade of B, C, or F with corresponding numerical values of 3.0, 2.0, and 1.0. An Azimuth/Elevation Error Composite, termed "Distance", was created to reflect the actual distance of the fired round from the target by first squaring the azimuth and elevation error scores for each engagement then taking the square root of the sum of those values.

The Tanks Crew Gunnery Skills Test (TCGST)

The content of the ET POI is driven by the Tank Crew Gunnery Skills Test (TCGST). The TCGST consists of 18 stations composed of various tank tasks that are scored on a GO/NO GO basis. These tests are administered as EOB tests at the completion of the relevant portion of the ET POI by the cadre who conduct the training. NTs are not normally trained or tested on the TCGST. However, to provide data for comparison of ETs and NTs on the ET training domain, the NT matches in our investigation were tested, without receiving any additional training beyond the NT POI, on 13 of the 18 stations. NTs were not tested by cadre on five TCGST stations because of limited available resources (e.g., tank time) and/or the safety risk created by having untrained personnel attempting difficult and dangerous tank tasks. Both ETs and NTs were tested via paper and pencil simulations during the Week 13 testing session on three of the five stations not tested by the cadre. Table 3 indicates which TCGST stations were cadre administered or tested via simulation by project personnel. Table 3 also identifies those stations judged by cadre personnel to be either too dangerous or too resource intensive to test both ET and NT soldiers.

Table 3. ET and NT TCGST Administration by Station

Station	Cadre Administered to ETs & NTs	Tested by Simulation Week 13	Safety Risk or Resources Too Great
1: ID Friendly & Threat Armored Vehicles	X		
2: ID & Explain Use 05-MM Main Gun Ammunition	X		
3: Clear, Disassemble, Perform Function Check, & Load 7.62-MM Coax Machine Gun	X		
4: Clear, Disassemble, Set Headspace & Timing, Perform Function Check, & Load Cal. .50 M2 HB Machine Gun	X		
5: Clear, Remove, Disassemble, Install, and Perform Function Check & Modified Firing Circuit Test on M68 Gun Breechblock		X	X
6: Boresight the 105-MM Main Gun	X		
7: Perform Replenisher Check	X		
8: Load 105-MM Main Gun	X		
9: Perform Failure-to-Fire Procedures on the 105-MM Main Gun	X		
10: Prepare Gunner's Station in M1 Tank for Operation	X		
11: Acquire Targets Through Thermal Imaging System (TIS)	X		
12: Engage Targets with 105-MM Main Gun from Gunner's Station in M1 Tank	X		

Table 3. ET and NT TCGST Administration by Station (continued)

Station	Cadre Administered to ETs & NTs	Tested by Simulation Week 13	Safety Risk or Resources Too Great
13: Prepare Tank Sketch Card	X		
14: Issue Initial and Subsequent Fire Commands		X	X
15: Estimate Range to Target		X	X
16: Prepare Tank for 3-Man Crew Operations & Fire Main Gun From TC Position			X
17: Lay Main Gun on Target			X
18: Mount, Adjust the Equilibrator, & Boresight Cal. .50 M2 HB Machine Gun with Commander's Weapon Sight	X		

Simulated TCGST Stations

As indicated, three of the five TCGST stations identified by the cadre as inappropriate for NT testing were tested via simulation during our Week 13 testing session. The TCGST stations simulated were Station 5: Remove, Disassemble, and Install the M68 Breechblock, Station 14: Issue Initial and Subsequent Fire Commands, and Station 16: Estimate and Determine Range to a Target. The Station 5 breechblock simulation used was that developed by Bessemer and Kraemer (1979). This test, originally developed for the M60A3 tank which uses the same M68 breechblock as the M1 tank, was modified for our purposes. A cadre SME knowledgeable of the M1 and A3 tanks identified items that were not appropriate for M1 crewmen because of differences in the installation and removal steps. Thirty-five of the 44 items on the original test were judged to be applicable to the M1 tank. The correct responses for the nine inappropriate items were marked on the answer sheet and subjects were told that those items would not affect their scores on the test. The breechblock simulation may be found in Appendix E (published in separate Research Note).

The Station 14 simulation, the Issue Initial and Subsequent Fire Commands Test, consisted of three battlefield scenarios presented pictorially and through a written description which was read aloud during the test session. The respondent was to write

the appropriate fire command on the answer sheet. Three cadre SMEs selected the scenarios to represent the type of fire commands tested in the TCGST from scenarios contained in the training materials prepared by Kraemer (1984). The SMEs also provided the correct fire commands for each selected scenario. The Fire Command simulation appears in Appendix F (published in separate Research Note).

The Station 15 simulation, the Range Determination Test, consisted of five multiple choice items, each consisting of a tank overlain by a gunner's primary sight reticle. Each item was selected from the Handbook for Sight Picture Training - M1 Tank (USARI, undated) by two cadre SMEs as depicting a situation similar to those for which the range determination is made in the field during TCGST testing. The range to the tank is determined on the simulation using knowledge of the reticle dimensions and the WORMS range computational formula. The Range Determination simulation is included in Appendix G (published in separate Research Note).

Concerns Regarding Army Administered TCGST Stations

The Army administered TCGST stations were subject to several of the same concerns identified for the Military Stakes and Gate III tests as well as several additional concerns. These included restricted score variance, scorer bias, and the reactive effects of testing. Our strategy for dealing with these issues follows.

Restricted Score Variance. Nine of the 18 TCGST stations are tested in the field where time and resources for testing are limited. Soldiers who fail require additional cadre time and resources for retraining and retesting. The situation was compounded by the fact that the testing of our NT soldiers in addition to the ET soldiers required nearly twice the normal testing time and resources. These constraints, as well as the influences identified for the Military Stakes and Gate III, operate to restrict the variability of scores on the TCGST.

Restricted score variance was dealt with by modifying the scoring procedure to include task proficiency ratings and the exact time to perform tasks. Existing scoring standards were reviewed and SMEs were interviewed to identify a meaningful, common underlying continuum to operationalize in the form of ratings. Task proficiency was determined to be the most relevant dimension across all tasks. Test administrators were consulted to gain a better understanding of commonalties underlying perceptions of task proficiency in order to develop the rating scale. A 5-point scale with behavioral anchors for the highest, middle, and lowest ratings was developed. Again, cadre SMEs assisted in the development of the scale and provided the anchor definitions. The score sheets for each TCGST station were modified to include the 5-point task proficiency scale.

Eight of the 13 TCGST stations administered to both ETs and NTs are required to be completed within a certain time limit. Cadre SMEs indicated that completing the task in less time typically reflected performance by a soldier who was more familiar with the proper procedure for task completion and subsequently demonstrated less hesitancy in his performance. Thus, less time to task completion indicates better performance. Since the exact time to task completion shows greater variability than the dichotomous GO/NO GO score (Morrison & Bessemer, 1981), we requested that the cadre collect the exact time for completion on these eight stations. Thus, the score sheets were further modified to include a place for recording the exact time for each task that had a time requirement.

In addition, three of the 13 TCGST stations (Station 1: Identify Friendly and Threat Armored Vehicles, Station 2: Identify and Explain the Use of the 105-MM Main Gun Ammunition, Station 13: Prepare Tank Sketch Card (also referred to as a range card)) provide continuous scores of the number correctly identified. These scores are normally recorded only as dichotomous GO/NO GO scores by TCGST test administrators. The continuous scores are likely to show greater variability among trainees. Thus, the score sheets for these three stations were further modified to include a place for recording the exact number of correct responses. The modified score sheets for each TCGST station appear in Appendix H (published in separate Research Note).

The cadre who administer the TCGST were trained in the use of the modified scoring procedure. A training session was held in which the objectives underlying the scoring modifications were explained; the cadre were instructed on the procedure for using the rating scale, the exact time measures, and the number correct measures; the rating scale and its anchors were discussed; and questions posed by the cadre were answered.

Scorer Bias. Scorer bias is an even more serious concern with the TCGST than it is with either the Military Stakes or Gate III. NT soldiers have not received training on many of the TCGST tasks. Thus, since the TCGST test administrators are also the trainers, they know which soldiers are ETs and which are NTs. Furthermore, safety standards dictate that test administrators know which soldiers have not received training on the tasks in order to prevent potential accidents. The Battalion was not able to comply with requests to have the TCGST administered to ETs and NTs EOC by "blind" administrators. Thus, during the training of the test administrators, it was stressed that knowledge of training track could result in scorer bias. Test administrators were instructed on techniques to try to minimize the impact of this potential confound.

Reactive Effects of Testing. Pilot administration of the TCGST revealed that NT soldiers felt somewhat frustrated and embarrassed to be repeatedly tested on tasks for which they received no training. This reactivity was addressed during data

collection by including a special briefing session during the administration of the initial Week 8 testing session used for collecting matching data. Subjects were told, among other things, that a sample of 1st Battalion OSUT soldiers would be tested on the TCGST, including NT soldiers who were not trained on the TCGST tasks. NT soldiers were told they should do their best to complete all tasks although they should expect to encounter some tasks they may be unable to complete. Comments from debriefing sessions at the end of data collection for subsequent cycles indicated that the pretesting briefing had the desired effect, that is, NT soldiers were more accepting of the TCGST testing, viewing it more as challenging than an humiliating.

TCGST Measures

There were a number of measures, both Army administered and ARI administered, collected for various TCGST stations. The measures for the Army-administered stations were first-round GO/NO GO (i.e., whether or not the soldier passed the station his first attempt), the exact time to complete the task for timed tasks, the exact number correct for those stations with continuous responses, and the task proficiency ratings for those stations to which it applied. The measure for the ARI-administered stations was the number correct for the given task. Table 4 details the specific measures obtained for each station.

Table 4. Measures Collected for Each TCGST Station

Station	GO/- NO GO	Task Proficiency	Exact Time	Number Correct
1	X	X		X
2	X	X		X
3	X	X	X	
4	X	X	X	
5*				X
6	X	X	X	
7	X	X	X	
8	X	X	X	
9	X	X	X	
10	X	X	X	
11	X	X	X	
12	X	X	X	
13	X			X
14*				X
15*				X
16				
17				
18	X	X	X	

* Tested via simulation in Week 13

TCGST data were collected for all five cycles of soldiers. However, data collection problems necessitated discarding the Army administered portion of the test for two cycles. Thus, we included TCGST data on the three simulated stations for all participants and data on the Army administered stations for soldiers in three of the five cycles.

TCGST Composites. As shown in Table 4, some 41 measures were collected across all TCGST stations by the Army and project staff. These measures represent both performance tests and paper and pencil tests. The three stations administered by the ARI project staff were paper and pencil simulations whereas the thirteen Army administered stations were performance tests. We formed two composites to reflect these distinctions, the "TCGST-ARI" composite and the "TCGST-Army" composite. Both composites are described in detail below.

The TCGST-ARI composite was formed by first converting each of the three measures from the ARI administered simulations (i.e., Stations 5, 14, & 15) to standard T scores. The TCGST-ARI composite is simply the average of the three standard scores from each of the paper and pencil simulations.

The formation of the TCGST-Army composite was somewhat more complicated. There were three stations for which fewer than 25 NTs and ETs were scored on our supplemental measures (i.e., Station 3 - exact time and proficiency rating; Station 4 - exact time; Station 13 - exact number identified). These data likely were not collected due to the previously identified constraints of testing NT soldiers in addition to the ET soldiers on the TCGST, i.e., limited time and resources. Our additional measures required administrative time beyond the simple GO/NO GO scoring. These four measures were dropped from further analyses. In addition, on two of these stations, Stations 3 and 4, there was no variance in the GO/NO GO scores, that is, all soldiers received a first-round GO on these stations. The GO/NO GO measures from Stations 3 and 4 were dropped from further analyses. Thus, in sum, all measures from Station 3 were eliminated from further analysis; the GO/NO GO measure and exact time measure were eliminated for Station 4, leaving only the task proficiency ratings to be analyzed for Station 4; and the exact number measure was eliminated for Station 13 leaving only the GO/NO GO measures to be analyzed for Station 13. Measures for the other Army administered stations stand as reported in Table 4.

The TCGST-Army composite was formed by computing a standard score for each of the twelve Army administered stations. This was accomplished by converting each valid measure for each station to Z-scores. These Z-scores for each station were then averaged to form a score for each station. Finally, the twelve station scores were averaged then converted to a T score to form the TCGST-Army composite.

In sum, the criterion space represented by the TCGST was reduced to two composite scores, one (TCGST-ARI) representing the three stations tested by ARI via paper and pencil simulation and the other (TCGST-Army) representing the 12 stations tested by the Army via hands-on performance tests. The correlation between these two composites is .26, indicating that the composites are measuring two different constructs.

Criterion Measures: NT-ET Common Domain

Project Alpha Army-Wide Ratings

Although the performance tests and paper and pencil knowledge test measure important aspects of the criterion space, there are other variables common to both the ET and NT training programs that are not reflected in these relatively brief tests. For example, it is difficult to develop a test that measures the amount of effort a soldier typically puts into his job. In order to measure the criterion space common to the ET and NT Training Programs reflecting typical performance, peer and supervisory ratings were obtained on the Project A Army-Wide Rating Scales. These scales are seven-point behaviorally anchored ratings scales tapping ten dimensions of performance, an Overall Effectiveness scale, and an NCO Potential scale. The ten dimensions are Technical Knowledge and Skill, Effort, Following Regulations and Orders, Integrity, Leadership, Maintaining Assigned Equipment, Military Appearance, Physical Fitness, Self-Development, and Self-Control. Each of the Army-wide scales is appropriate for rating entry level performance in any Army MOS.

Rater Assignment

Four peer-raters were assigned to rate each participant. The raters were assigned for each cycle of soldiers according to the process described in the Project A protocol (Administrator's Manual: Peer and Supervisor Rating Sessions for Concurrent Validation June - November 1985, 1985). One week prior to the testing session in which the ratings were collected, each ET or NT soldier was asked to indicate the five peers with whose performance they were most familiar and whom they could most accurately rate. They were also asked to indicate other peers that may be slightly less familiar but for whom they could still provide accurate ratings. Those raters who indicated they could evaluate the fewest number of peers were assigned ratees first. Through several iterations of assignments, a minimum of four peer raters was were assigned for each ratee and no rater was required to evaluate more than four peers.

Supervisors were likewise assigned ET/NT-ratees according to the Project A guidelines for determining raters. A cadre SME was asked to list at least two NCOs familiar enough with each soldier's performance to provide ratings for that soldier. Each ratee was assigned to be evaluated by two NCOs and each NCO rater rated no more than ten soldiers.

Rating Sessions

The peer ratings were collected during the Week 13 testing session. Extensive instructions were provided to the raters. These instructions included the Project A Rater Training Program (Administrator's Manual: Peer and Supervisor Rating Sessions for Concurrent Validation June - November 1985, 1985) explaining common rating errors and how to avoid them. Supervisory ratings were obtained in a session held for the NCO raters during the final week of each cycle. Supervisory raters also received the Project A Rater Training. The rating task was not timed, that is, both peer and supervisor raters were allowed as much time as needed to complete all ratings.

It should be noted that most raters, peers and NCOs, were aware which soldiers were in the ET and NT training tracks. This source of contamination in the ratings was difficult to avoid in our situation. There are, however, several factors that might somewhat mitigate the contaminant. The peer ratings were embedded in the Week 13 testing session. Soldiers were not aware that this evaluation session was for the purpose of differentiating ET and NT soldiers and should not have been focusing on the ET-NT distinction as they made their ratings. Although the NCO raters were not specifically informed of the purpose of the ratings, it is likely that they suspected the objective behind them. When other cadre members were questioned about biasing effects, they indicated that NCOs had mixed feelings about the ET program, that is, some were very much in favor of it while others were against it. Generally those NCOs against the ET program feel that it results in an "instant NCO" who doesn't have the requisite skills and experience acquired through a slower progression through the ranks. Assuming that both views were equally represented in our raters, their attitudes may not systematically bias the ratings in favor of or against ETs. These considerations paired with the careful assignment of multiple raters for each ratee and the administration of the rater training program hopefully minimized the knowledge of training track contaminant.

Project A Army-Wide Rating Measures

Project A Army-Wide ratings were obtained for each ET and NT soldier on ten performance dimensions, on an Overall Effectiveness scale, and on an NCO Potential scale from four peer raters and from two NCO raters. The ratings for each of the twelve scales were averaged across the four peer raters to obtain a single "peer rating" and across both NCO raters to obtain a single "NCO" rating.

Analyses and Results

The previous section described the development of a large and comprehensive set of criterion measures for comparing ET/NT performance. Not only were measures developed which sampled differentially from the content of the two training tracks, but multiple assessment methods were utilized. The result is a series of performance measures derived from hands-on, paper-and - pencil, and supervisor/peer ratings.

Testing the significance of obtained ET/NT performance differences on every performance measure is unwise. Aside from varying degrees of redundancy in various subsets of our criterion measures, a large number of tests of significance on data obtained from our sample alone, increases the likelihood of detecting differences which do not exist in the population. Accordingly, before addressing the differences in ET/NT performance, we began our analysis of each set of criterion measures by considering strategies for combining measures to form a smaller number of meaningful composites. In the following sections, we discuss each set of criterion measures in turn. First we describe the strategies used to form composites and the nature of the resulting composites. This is followed by a description of our analyses of ET/NT performance. Each section concludes with a description of the results of the analyses.

UCOFT Criterion Refinement

Though nine performance measures were obtained for each exercise, these measures can be rationally classified into three components: aiming accuracy, response latency, and system management. Accuracy measures include hit rate, reticle aim, elevation, azimuth, and distance. All basically reflect the resulting proximity of the projectile to the target. Identification time, opening time, and target acquisition all provide information about latency, that is how quickly the targets are identified and fired upon. System management indicates the appropriateness of switch settings and ammunition selection. With an eye toward forming the above three composites, our analyses of the UCOFT measures began by examining the intercorrelations among the nine performance indicators (Table 5) as well as their test-retest reliabilities (Table 6).

Table 5. Intercorrelations Among Nine UCOFT MEASURES

	IDTIME	OPEN	HIT	TA	SM	RA	AZ	EL	DIST
IDTIME	1.0000	.5028**	-.4993**	-.9317**	-.3737**	-.5451**	.3881**	.4263**	.4199**
OPEN	.5028**	1.0000	-.4559**	-.5943**	-.7600**	-.6659**	.3328**	.2377*	.2364*
HIT	-.4993**	-.4559**	1.0000	.6047**	.5749**	.8881**	-.5862**	-.7128**	-.6111**
TA	-.9317**	-.5943**	.6047**	1.0000	.5305**	.6498**	-.4851**	-.4354**	-.4258**
SM	-.3737**	-.7600**	.5749**	.5305**	1.0000	.6998**	-.4084**	-.3793**	-.3005**
RA	-.5451**	-.6659**	.8881**	.6498**	.6998**	1.0000	-.5257**	-.6580**	-.5369**
AZ	.3881**	.3328**	-.5862**	-.4851**	-.4084**	-.5257**	1.0000	.3415**	.8789**
EL	.4263**	.2377*	-.7128**	-.4354**	-.3793**	-.6580**	.3415**	1.0000	.6030**
DIST	.4199**	.2364*	-.6111**	-.4258**	-.3005**	-.5369**	.8789**	.6030**	1.0000

N of cases: 165 1-tailed Signif: $^{*}p<.01$ $^{**}p<.001$

With respect to the accuracy measures, inspection of Table 6 reveals that the distance measure is unreliable (rel = .35). This consideration, together with the fact that distance is simply a composite of two other measures in the accuracy group and thus contains no unique information, argued for dropping this measure from further analyses. Reticle aim was also dropped. Though it does possess acceptable reliability, this measure is itself a composite of opening time, kill time, and azimuth and elevation errors. Thus it is largely redundant with measures comprising the accuracy composite, and to a lesser extent overlaps with the latency composite. The lack of unique information provided by reticle aim is also revealed by the multiple correlation (R = .89) between reticle aim on the one hand and the three composites of accuracy, latency, and system management. Thus, the accuracy composite was composed of the hit rate, azimuth, and elevation measures.

Table 6. UCOFT Measures Reliability Coefficients

UCOFT Measure	Uncorrected Reliability	Corrected Reliability
Target Acq.	.73	.84
System Mgt.	.47	.64
Reticle Aim	.47	.64
Hit Rate	.40	.57
Target ID Time	.78	.88
Opening Time	.69	.82
Azimuth Error	.27	.43
Elevation Err.	.35	.52
Distance	.21	.35

Having decided on the components of the accuracy composite, the manner of combining these measures was addressed. In order to increase the reliability of the composite, each measure was converted to a z score, then multiplied by its reliability before

summing the measures. Since low scores on azimuth and elevation are desirable, these measures were reflected before summing. Finally, the resulting vector of accuracy scores was converted to T scores.

Identification time, opening time and target acquisition were used to form the latency composite. All components have reliability coefficients above 0.80. Though target acquisition is highly correlated with identification time, the measure was nevertheless retained. Our rationale for doing so after dropping reticle aim in part for a similar reason is that, unlike reticle aim which is a mixture of time and accuracy metrics, target acquisition is largely a speed measure. As with accuracy, the latency composite was formed by weighting the standardized components by their reliabilities before summing the reflected identification time and opening time scores with the target acquisition score. The latency composite scores were then converted to T scores.

The third measure, system management, though itself a composite, was not combined with any other measures. The only modification of the system management score was to convert it to a T score to facilitate comparisons with the other measures.

The test-retest reliability of the accuracy and latency composites was assessed by computing the composites twice, once for each set of six exercises, and then correlating the two scores from each set of exercises. The resulting coefficients were then corrected using the Spearman-Brown formula. The results of these calculations are shown in Table 7.

Table 7. UCOFT Composite Reliability Coefficients

UCOFT Composite	Uncorrected Reliability	Corrected Reliability
Accuracy	.41	.58
Latency	.76	.86
System Mgmt.	.47	.64

Though the Latency composite demonstrates good reliability, both Accuracy and System Management are lower than is desirable. Still, these reliabilities are adequate for our purposes; they are in the range typically found for rating criteria. Clearly, improving reliability by increasing the number of exercises over which these composites are computed is indicated in future efforts where sufficient UCOFT time can be allocated. Attenuated reliabilities on the criterion measures reduce the probability of detecting ET/NT performance differences where they exist.

UCOFT Composite ET/NT Performance Comparisons

Examination of the intercorrelations among the three UCOFT composites (Table 8) reveals that although each measure provides unique information, the measures are by no means independent. Consequently, we began our analysis of ET/NT performance by performing a multivariate analysis of variance (MANOVA) on the three composites. In general, follow-up univariate analysis of the individual composites is appropriate only if the multivariate test of an effect is significant (Bernstein, 1988).

Table 8. UCOFT Composite Intercorrelations

	ACCURACY	LATENCY	SYS.MGMT.
LATENCY	.5807**	1.0000	.6259**
SYS.MGMT.	.5437**	.6259**	1.0000

The design can be viewed as a randomized block MANOVA with training track (Training) as the treatment and tank commander (TC) as the blocking variable. Results of the MANOVA are shown in Table 9. The tests for the multivariate effects are displayed first. For all multivariate analyses the Pillai-Bartlett test statistic was used to test for multivariate effects (Norusis, SPSS/PC+, 1986).

Table 9. UCOFT Composites Multivariate Significance Tests

Effect	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
TC	.29782	9.21465	6.00	316.00	.000
Training	.05372	2.97105	3.00	157.00	.034
Trng. x TC	.04377	1.17830	6.00	316.00	.318

Significant main effects were obtained for both TC and training; their interaction was not significant. The significant TC effect is of little interest here since TC effects are controlled by our design. However, the TC effect further underscores the need to control for TC in UCOFT studies of gunner performance. Despite the intensive TC training designed to standardize their performance and despite analyses of tape recordings of TC commands during data collection which revealed high TC accuracy and consistency, Table 10 shows that all three composites were affected by TC performance.

Table 10. Univariate Tests of TC Effects on UCOFT Composites

Variable	Df.	Hyp.MS	Err.MS	F	Sig
ACCURACY	2,159	430.80	91.92	4.68	.011
LATENCY	2,159	2229.81	73.76	30.22	.000
SMT	2,159	608.70	92.74	6.56	.002

Of primary interest is the impact of training track on UCOFT performance. Given the significant training effect in the MANOVA, univariate significance tests were computed to evaluate the impact of training on each of the UCOFT composites. Table 11 indicates that accuracy and system management were affected by training; latency was not.

Table 11. Univariate Tests of Training Effects on UCOFT Composites

Variable	Df.	Hyp.MS	Err.MS	F	Sig
ACCURACY	1,159	732.26	91.92	4.68	.005
LATENCY	1,159	98.66	73.76	30.22	.249
SMT	1,159	359.97	92.74	6.56	.050

ET soldiers were more accurate and made fewer system management errors than their NT cohorts. Table 12 shows that ETs outperformed NTs by a third of a standard deviation or more on these two measures. Considering the reliability of these composites, the true differences between the training track means is in all likelihood greater than reported. Excellence track training does result in enhanced UCOFT performance.

Table 12. UCOFT Composite Means & SDs By Training Track

	TRAINING	
UCOFT COMPOSITE	ET	NT
Accuracy		
Mean	51.94	48.04
Std. Dev.	9.38	10.27
Latency		
Mean	50.87	49.11
Std. Dev.	9.90	10.09
System Mgmt.		
Mean	51.58	48.40
Std. Dev.	10.31	9.47

The Impact of U-COFT Engagement Degradation on ET vs. NT Performance.

Above we reported superior performance of ETs on two of three measures of UCOFT performance. It is reasonable to hypothesize that the enhanced performance exhibited by ETs stems from their capacity to better adjust to exercises where one or more tank systems are inoperative. Under these "non-standard" conditions, we might expect the better conceptual, theory-based training of ETs to manifest itself. Here, human skills must replace the more automated fire control systems built into the M1. Moreover, under actual battle conditions, it is perhaps more realistic to assume that the achievement of mission objectives will turn on the tank crew's ability to fire effectively under these "degraded" conditions.

The present data base affords us an opportunity to begin to look at this question. Of the six UCOFT exercises, two require the gunner to respond to targets where all systems are functioning properly (see Table 2 - UCOFT Exercises 322230 & 325120). The remaining four exercises require the gunner to perform where one or more fire control systems are malfunctioning. The nature of the malfunctions is detailed in Table 2.

To examine the impact of engagement mode ("normal" vs. degraded) on the relative performance of ETs vs. NTs, we twice computed the three composites described above. One set of the three composites was derived from performance on the two normal exercises, the other set was calculated based on performance on the four degraded exercises. These two sets of composites were then analyzed for ET/NT differences.

Before describing these analyses, two potential confounds must be mentioned: differential practice effects and differential composite reliability. Since the normal/degraded issue and resulting analyses emerged after the data were collected, neither the proportional mixture nor the sequencing of exercises was guided by design considerations dictated by these research questions.

Under the assumption that UCOFT performance will improve with practice, the practice effect problem arises from the juxtaposition of the normal and degraded exercises within the six exercise sets. Were the normal engagements presented first, then practice would serve to lessen the impact of fire control system malfunctions on performance. Conversely, presenting the degraded scenarios first would lead to overstating the impact of system malfunctions. Quite fortuitously, the normal mode exercises were presented first and last, with the four degraded scenarios in between. Given this sequencing, we believe practice effects are not a serious confound.

Differences in composite reliability for the normal compared to the degraded composites is a potentially more serious problem. Because there are twice as many degraded as normal exercises, the composites based on the degraded scenarios are likely to be more reliable. Table 13 below displays the estimated normal and degraded composite reliabilities calculated for the three composites by stepping down the six-exercise based reliabilities reported in Table 7.

Table 13. UCOFT Normal/Degraded Composite Reliability Coefficients

UCOFT Composite	Normal	Degraded	Combined
Accuracy	.32	.48	.58
Latency	.68	.81	.86
System Mgmt.	.38	.55	.64

Obviously, both the normal and the degraded mode engagement composite reliabilities are low. This will only serve to mask any true differences that may exist for either condition. The fact that the normal composite reliabilities are systematically lower than the degraded mode suggests that this making effect is likely to be greater for the normal mode comparisons. That is, ET/NT differences are going to be more difficult to detect for these exercises. While this effect cannot be teased from the present data, awareness of the problem should facilitate our interpretation of the results which follow.

The normal and degraded UCOFT composites described above were analyzed using a repeated measures (normal vs. degraded) split-plot MANOVA (Norusis, SPSS/PC+, 1986). As before, the training track and tank commander are between factors; however now engagement mode is added to to the design as a within subjects factor. The results of the MANOVA are shown in Table 14.

The multivariate tests on the between subjects effects are essentially the same comparisons presented in Table 9. No new information is provided. Though the results generally parallel those reported in Table 9, the reader will note that the significance level for the training main effect is .064 here rather than .034 reported in Table 9. This discrepancy arises from the reduction in composite reliabilities employed in this analyses (see Table 13). It simply illustrates the point raised earlier regarding the masking effect of lower composite reliabilities on our design's sensitivity to training track differences.

Table 14. UCOFT Composites Multivariate Significance Tests: Normal vs. Degraded Exercises

Effect	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Between...					
TC	.28999	8.93143	6.00	316.00	.000
Training	.04514	2.47444	3.00	157.00	.064
Trng. x TC	.04530	1.22050	6.00	316.00	.295
Within...					
Mode	.00058	.03038	3.00	157.00	.993
TC x Mode	.01356	.35961	6.00	316.00	.904
Trng x Mode	.05891	3.27590	3.00	157.00	.023
Trng x TC x Mode	.01456	.38631	6.00	316.00	.888

The test for the within subjects mode main effect is neither meaningful nor of particular interest. The reader will recall the degraded and normal mode UCOFT measures were each standardized to ensure appropriate weighting for the variables forming each composite. The mode main effect was thereby eliminated. Since this main effect is of little interest, the increased meaningfulness of the composites resulting from this procedure justifies the loss of information about this effect.

Table 14 shows a significant effect for the training by mode interaction. This is the effect of primary interest, since it addresses the question of whether mode moderates the effect of training on performance. Clearly it does. Given the significant interaction for the multivariate test, we examined the univariate tests on each of three UCOFT composites. Table 15 presents these results.

Table 15. Univariate Tests of Training by Mode Interaction Effects on UCOFT Composites

Variable	Df.	Hyp.MS	Err.MS	F	Sig
ACCURACY	1,159	208.11	64.08	3.25	.073
LATENCY	1,159	5.17	23.40	.22	.639
SMT	1,159	355.79	53.92	6.60	.011

The univariate tests reveal the training by mode interaction is present for system management, and assuming a one-tailed test is appropriate here, the interaction also occurs for accuracy. A plot of the interaction for accuracy and system management are presented in Figures 1 and 2. Table 16 displays the means and

standard deviations for the three composites. As hypothesized, the superiority of ET over NT performance is manifest in the degraded exercises; there is no discernible difference in performance under normal conditions.

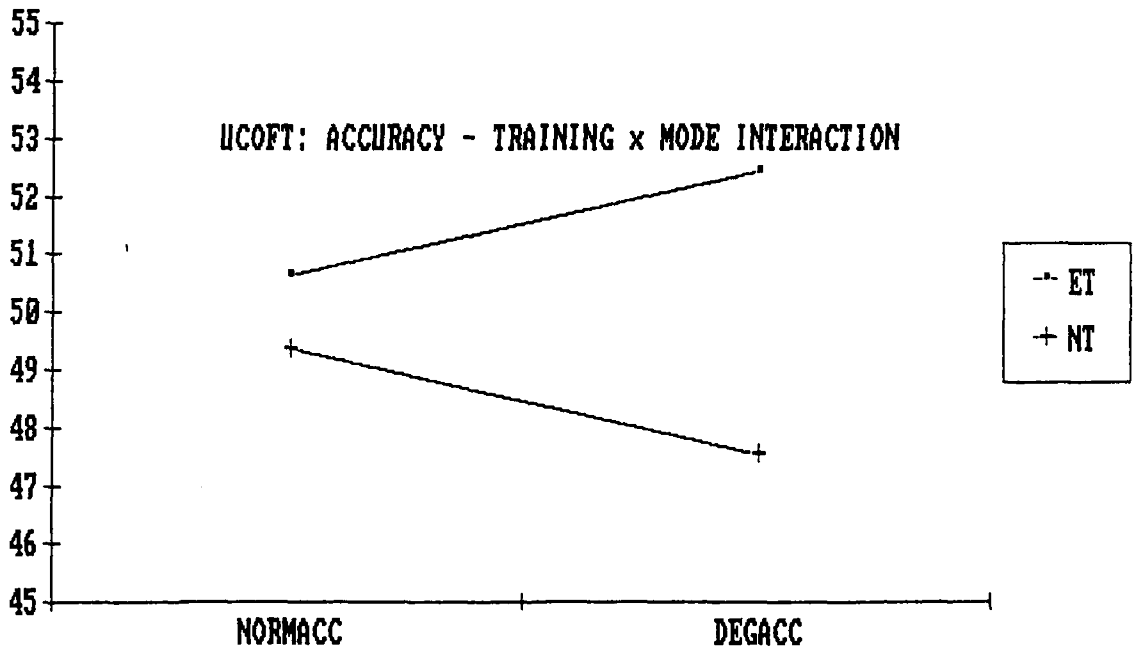

Figure 1. UCOFT accuracy training by mode interaction.

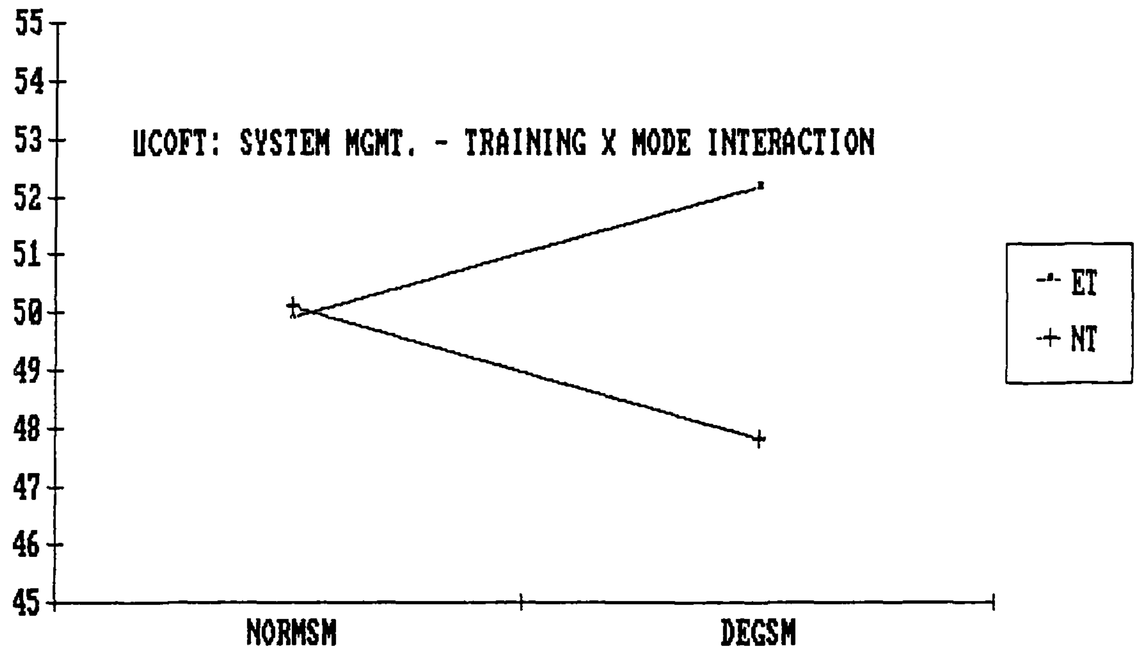

Figure 2. UCOFT systems management training by mode interaction.

Table 16. *UCOFT Composite Means & SDs By Training Track & Mode*

UCOFT COMPOSITE	NORMAL		DEGRADED	
	ET	NT	ET	NT
Accuracy				
Mean	50.62	49.36	52.42	47.55
Std. Dev.	9.46	10.54	9.50	9.95
Latency				
Mean	50.92	49.07	50.76	49.23
Std. Dev.	9.82	10.15	10.05	9.96
System Mgmt.				
Mean	49.91	50.91	52.16	47.81
Std. Dev.	9.94	10.12	10.04	9.52

In summary, in terms of performance on the UCOFT, the ET program clearly results in measurable gains over the performance achieved by NT soldiers. Gunner accuracy and system management are handled better by ETs, with no difference between ETs and NTs in response latency. The superiority of ETs on these measures can be traced directly to their greater skill in handling gunner tasks when one or more fire control systems is disabled.

TCGST ET/NT Composite Performance Comparisons

Earlier we described two composites developed to summarize performance on the rather large number of TCGST measures. These composites are labeled "TCGST-ARI" and "TCGST-Army" in the following analyses. The TCGST-ARI is a composite of paper and pencil tests developed especially for this evaluation effort and collected by project staff; the TCGST-Army is a composite of hands-on performance measures traditionally used in TCGST evaluations and was collected by Army personnel.

The reader will recall that out of concern for the integrity of the scoring process used in collecting the first two cycles of the measures forming the TCGST-Army composite, these measures from these cycles were dropped. Consequently, TCGST-ARI is available on 166 soldiers, but TCGST-Army is available on only the 92 soldiers from the last three cycles. To ensure that ET/NT performance differences were not systematically biased in this reduced sample, we examined the difference in ET/NT performance on TCGST-ARI as a function of training track and whether or not we had TCGST-Army. Using the TCGST-ARI as the dependent variable, a two-way ANOVA revealed significant training effects,

significant effects due to whether or not we have TCGST-Army, but fortunately, no significant interaction. Scores on TCGST-ARI are higher for the soldiers from the first two cycles, but the difference in ET/NT performance is not significantly different for these cycles compared to the last three cycles. Since our primary interest is the difference in ET/NT performance, these results suggest that the missing data, although reducing power, should not distort our findings with respect to the primary research questions.

Table 17 displays the results of the MANOVA examining the impact of training program on the two TCGST composites. Given the significance of the resulting test statistic, separate univariate ANOVAs were performed on each TCGST composite. These results appear in Table 18.

Table 17. TCGST Criteria MANOVA Summary Table

Effect	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Training	.61831	72.08597	2.00	89.00	.000

Table 18. Univariate Tests of TCGST Criteria

Variable	Df.	Hyp.MS	Err.MS	F	Sig
TCGST-ARI	1,90	744.79	92.58	8.04	.006
TCGST-Army	1,90	5561.90	39.31	141.48	.000

Significant ET/NT differences exist for both the ARI and the Army composites. From Table 19 it is apparent that ETs perform better on both the hands-on TCGST performance composite and the paper and pencil TCGST measure. On the hands-on measure it is noteworthy that not only is the NT performance lower, but the standard deviations indicate that NT performance is far more variable. This undoubtedly results from the lack of training NTs received on the TCGST measures.

Table 19. TCGST Composite Means & SDs By Training Track & Mode

TCGST COMPOSITES	ET	
	ET	NT
ARI MEASURES		
Mean	51.31	45.62
Std. Dev.	9.50	9.74
ARMY MEASURES		
Mean	57.78	42.23
Std. Dev.	4.07	7.88

In sum, the TCGST composites document performance gains for ETs over their NT cohorts. Since the TCGST samples from the ET training content domain, this finding offers support for the ET program.

Military Stakes ET/NT Performance Comparisons

This section describes the analyses and results for the Military Stakes criteria and the NT paper and pencil measure (NTPP). These measures were designed to sample from the NT training content domain. Given our hypothesis that compression does not result in a decrement in ET performance relative to NT performance, we are essentially attempting here to confirm the null hypothesis!

As described earlier, six measures were derived from the Military Stakes testing process: time to complete the Military Stakes course, pistol maintenance, range estimation, rifle maintenance, identify friendly/threat vehicles, and identify threat aircraft. Although the NTPP is not a part of the Military Stakes, since it was also developed to sample the NT domain it is included in the analyses of the Military Stakes measures. In this way we limit somewhat, the number of separate significance tests that are performed.

The effect of training track on the six Military Stakes criteria and the one NTPP test score was evaluated by computing a simple one-way MANOVA. Only the 124 soldiers, 64 ETs and 60 NTs, for whom complete data on the seven measures were available were included in this analysis. Table 20 reveals that significant differences do exist between the ET/NT mean vectors.

Table 20. Military Stakes Criteria MANOVA Summary Table

Effect	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Training	.17550	3.57243	7.00	116.00	.002

The results of the follow-up univariate analyses are displayed in Table 21. Table 22 shows the means and standard deviations by training track for each criterion measure. Though no differences were hypothesized, significant training track differences were found for identifying friendly/threat vehicles and for the NTPP.

Table 21. Univariate Tests of Military Stakes Criteria

Variable	Df.	Hyp.MS	Err.MS	F	Sig
TIME	1,122	293.57	102.88	2.85	.278
PISTOL	1,122	119.89	100.77	1.19	.278
RANGE	1,122	38.70	99.88	.39	.535
RIFLE	1,122	205.50	114.04	1.80	.182
ID. VEH.	1,122	970.27	101.46	9.56	.002
ID. AIR.	1,122	252.72	114.60	2.20	.140
NTPP	1,122	428.84	91.22	4.70	.032

Inspection of the means in Table 22 reveals that with respect to both significant effects, ETs perform better than NTs. Thus the impact of compression notwithstanding, ETs perform as well or better than NTs on measures sampling from that portion of the training content domains on which NTs spend the greater amount of training time.

Table 22. Military Stakes Criteria Means & SDs By Training Track

Military Stakes Criterion	ET	
	ET	NT
ID VEHICLE		
Mean	52.50	47.52
Std. Dev.	8.98	10.45
ID AIRCRAFT		
Mean	48.76	51.21
Std. Dev.	10.47	9.36
TIME		
Mean	51.82	48.12
Std. Dev.	10.45	9.17
PISTOL		
Mean	48.57	51.45
Std. Dev.	11.29	8.43
RANGE		
Mean	50.62	49.41
Std. Dev.	9.54	10.41
RIFLE		
Mean	49.12	50.93
Std. Dev.	11.20	8.54
NTPP		
Mean	52.09	47.79
Std. Dev.	9.67	9.98

Performance Ratings ET/NT Comparisons

The results of the analyses of both the supervisor and the peer ratings are presented in this section. The analyses of each set of ratings proceeded in the same manner. First, the mean ratings across raters on each of the twelve Army-wide performance dimensions was subjected to a MANOVA. Table 23 and Table 24 display the results for supervisor ratings and peer ratings, respectively. Both MANOVAs argued for the rejection of the hypothesis of no differences in the performance rating mean vectors between ETs and NTs. Accordingly, follow-up univariate ANOVAs were performed.

Table 23. Supervisory Ratings Criteria MANOVA Summary Table

Effect	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Training	.31319	5.58600	12.00	147.00	.000

Table 24. Peer Ratings Criteria MANOVA Summary Table

Effect	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Training	.36131	7.21274	12.00	153.00	.000

The results of the univariate tests for mean differences between supervisor ET/NT performance ratings on each of the rating scales is presented in Table 25. The parallel results for peers are shown in Table 26. All twelve scale means reveal significant ET/NT differences. This is true for supervisor as well as peer ratings

Table 25. Univariate Tests of Supervisor Ratings Criteria

Rating Scale	Df.	Hyp.MS	Err.MS	F	Sig
Technical Knowledge/Skill	1,158	28.92	.95	30.15	.000
Effort	1,158	21.44	1.51	14.15	.000
Following Regs./Orders	1,158	13.48	1.39	9.69	.002
Integrity	1,158	26.68	1.47	18.13	.000
Leadership	1,158	53.36	1.71	31.03	.000
Equipment Maintenance	1,158	14.28	1.16	12.31	.001
Military Appearance	1,158	22.64	1.03	21.82	.000
Physical Fitness	1,158	35.67	1.18	30.06	.000
Self-Development	1,158	31.46	1.17	26.88	.000
Self-Control	1,158	20.59	1.30	15.83	.000
Overall Effectiveness	1,158	41.03	.99	41.44	.000
NCO Potential	1,158	67.66	1.55	43.50	.000

Table 26. Univariate Tests of Peer Ratings Criteria

Rating Scale	Df.	Hyp.MS	Err.MS	F	Sig
Technical Knowledge/Skill	1,164	41.00	.63	65.25	.000
Effort	1,164	21.93	.79	27.78	.000
Following Regs./Orders	1,164	30.37	.99	30.63	.000
Integrity	1,164	19.92	.89	22.40	.000
Leadership	1,164	52.00	1.14	45.40	.000
Equipment Maintenance	1,164	23.59	.63	37.68	.000
Military Appearance	1,164	19.57	.80	24.57	.000
Physical Fitness	1,164	31.88	1.07	29.80	.000
Self-Development	1,164	32.69	.71	45.83	.000
Self-Control	1,164	7.70	1.52	5.06	.026
Overall Effectiveness	1,164	28.34	.61	46.64	.000
NCO Potential	1,164	57.56	1.06	45.24	.000

Inspection of the actual scale means in Table 27 and in Table 28 shows that ETs consistently receive higher ratings than their NT cohorts, regardless of whether the raters are supervisors or peers. In most instances the differences are substantial, typically a full standard deviation higher for ETs. It is clear that ETs are generally perceived as better performers than NTs across a wide variety of performance dimensions.

Table 27. Supervisor Ratings Means & SDs By Training Track

SUPERVISOR RATINGS	TRAINING	
	ET	NT
Tech. Knowledge/Skill		
Mean	5.44	4.59
Std. Dev.	.98	.97
Effort		
Mean	5.17	4.43
Std. Dev.	1.21	1.25
Following Regs./Orders		
Mean	5.28	4.69
Std. Dev.	1.11	1.24
Integrity		
Mean	5.34	4.51
Std. Dev.	1.17	1.25
Leadership		
Mean	5.28	4.13
Std. Dev.	1.25	1.36
Equipment Maintenance		
Mean	5.23	4.61
Std. Dev.	1.08	1.09
Military Appearance		
Mean	5.48	4.72
Std. Dev.	.89	1.12
Physical Fitness		
Mean	5.49	4.56
Std. Dev.	.90	1.24
Self-Development		
Mean	5.43	4.54
Standard Deviation	.97	1.17
Self-Control		
Mean	5.58	4.85
Std. Dev.	.91	1.32
Overall Effectiveness		
Mean	5.53	4.51
Std. Dev.	.91	1.06
NCO Potential		
Mean	5.43	4.13
Std. Dev.	1.05	1.41

Table 28. Peer Ratings Means & SDs By Training Track

PEER RATINGS	TRAINING	
	ET	NT
Tech. Knowledge/Skill		
Mean	5.05	4.05
Std. Dev.	.70	.87
Effort		
Mean	4.45	3.73
Std. Dev.	.70	1.04
Following Regs./Orders		
Mean	4.95	4.10
Std. Dev.	.91	1.08
Integrity		
Mean	4.70	4.01
Std. Dev.	.88	1.00
Leadership		
Mean	4.56	3.44
Std. Dev.	1.04	1.10
Equipment Maintenance		
Mean	4.90	4.14
Std. Dev.	.74	.84
Military Appearance		
Mean	5.07	4.38
Std. Dev.	.81	.97
Physical Fitness		
Mean	5.18	4.30
Std. Dev.	.88	1.17
Self-Development		
Mean	4.80	3.92
Standard Deviation	.76	.92
Self-Control		
Mean	4.63	4.20
Std. Dev.	1.16	1.30
Overall Effectiveness		
Mean	5.05	4.22
Std. Dev.	.69	.86
NCO Potential		
Mean	5.44	4.59
Std. Dev.	.98	.97

DEVELOPING AND COMPARING ET, NT, AND ANCOC SOLDIERS' APTITUDE, INTEREST, AND TEMPERAMENT PROFILES

The Army has an interest in retaining its better enlistees and encouraging their advancement to NCO ranks. Presumably, selection as an ET suggests that during this early point in the soldier's career, the soldier has been identified as a likely candidate for eventual NCO status. Conversely, persons not selected for ET have been judged somewhat less likely candidates for eventual promotion to NCO. It thus seems reasonable to assume that the aptitude, interest, and temperament profile of ETs might more closely resemble the profile of current NCOs than do the profiles of NTs. If this is so, OSUT soldier profiles might provide an additional source of guidance to the Army in the early identification of NCO prospects and in channeling these soldiers into the ET program. The present data base provides an opportunity to take a preliminary look at this potential value of these profiles. Thus, in this section we describe our strategy for accomplishing the third objective of this project, to evaluate the degree of similarity between the aptitude, interest, and temperament profiles of ET, NT, and ANCOC soldiers.

Selection of Aptitude, Interest, and Temperament Constructs and Measures

The literature previously cited demonstrates the potential predictive value of a variety of cognitive, perceptual, psychomotor, and biodata measures for both tank commanders and gunners. Our objective was to identify a set of measures that differentiates ETs, NTs, and senior NCOs. We believed a firm rationale existed for using the ARI Project A Predictor Battery (PAPB) together with ASVAB composites. This belief was based primarily on the conceptual framework and the care which guided the development of the PAPB and the traditional use of the ASVAB by the Army.

PAPB Measures

A driving force which guided the PAPB development was expansion of the predictor domain beyond the cognitive abilities already successfully tapped by the ASVAB. In particular, rapid advances in the capabilities and availability of microcomputers has enabled the reliable measurement of a greater variety of psychomotor and perceptual constructs. The PAPB includes computer-based measures of ten abilities largely independent of ASVAB content. These are 1) simple reaction time, 2) choice reaction time, 3) one-handed tracking, 4) two-handed tracking, 5) target shoot-tracking, 6) a target identification measure of perceptual speed, 7) a cannon shoot measure of perceptual judgment, 8) perceptual speed & accuracy, 9) a number memory/tracking task, and 10) a measure of short-term memory.

Several of these constructs appear strikingly similar to those underlying the device-dependent work sample measures validated in several tank gunner studies (Campbell and Black, 1982; Biers and Sauer, 1982). Preliminary analyses of these PAPB measures reveals that fifty to eighty percent of their reliable variance is uncorrelated with ASVAB scores (McHenry, J. J. Personal communication. November 1985). Thus they offered considerable hope for capturing criterion variance unexplained by the ASVAB.

The PAPB also includes paper and pencil measures of constructs unexplored in the ASVAB. The Assessment of Background and Life Experiences (ABLE) taps a variety of temperament dimensions, the Army Vocational Interest Career Examination (AVOICE) contains a number of interest and biodata scales. In addition, the paper and pencil battery includes measures of spatial visualization, spatial orientation, and induction (figural reasoning). Though little attention has been devoted to evaluating the utility of these types of measures for predicting tank crewman success, limited evidence suggests at least biodata can produce some validity over the ASVAB (Campbell and Black, 1982).

A potentially important variable that differentiates ET and NT soldiers is motivation. Reportedly, ET soldiers are highly motivated to perform well throughout the course of training. It was therefore of interest to include some measure of motivation in the predictor space to determine if indeed there is a difference between ET, NT, and ANCOC soldiers on this variable. The PAPB contains a number of scales on the ABLE and AVOICE that proved useful in this regard.

There were three additional attractive aspects of the PAPB. Substantial progress had already been made in evaluating the psychometric characteristics of the PAPB. Reliability and factor analytic studies suggest most of the scales possess adequate stability and are not highly inter-correlated (Hough, Barge, Houston, McGue, & Kamp, 1985; Toquam, Dunnette, Corpe, McHenry, Keyes, McGue, Houston, Russell, & Hanson, 1985). Studies of a practice effect on the psychomotor tests indicate very little score variance is attributable to this potential contaminant (J. McHenry, personal communication, December, 1985). Finally, administration time is also quite reasonable given the comprehensive nature of the battery. Both the paper and pencil and the computer-based tests were easily administered in three hours, each component requiring approximately 1-1/2 hours.

Reduction of the Predictor Space

It was necessary to limit the number of profile dimensions to between ten and fifteen given the number of available ANCOC's, ETs, and NTs. Taken together, there are in excess of fifty scales on the ASVAB and PAPB. This number does not include

composites derived from various combinations of these scales. We relied on expert judgment to select the PAPB scales to use in the profile comparison. The available Project A measures included six psychomotor scores from the computer battery and more than twenty scales from the ABLE and AVOICE.

Selection of Psychomotor Measures

Based on the recommendation of Dr. Jeff McHenry (personal communication, March 1986), we decided to use Factor 1 and Factor 2 from the PAPB computer battery. Project A analyses of the psychomotor test revealed three factors: Factor 1, the mean of the log of the distance score for the two tracking tests; Factor 2, the composite mean of the distance score from the target shoot and time discrepancy score from the cannon shoot; Factor 3, the composite of mean target shoot time to fire score and mean movement time across all reaction tests (J. McHenry, personal communication, March 1986). The three factors are listed in the order of the variance accounted for. In addition, as shown in Table 29 below, the tests comprising Factor 1 are the most reliable (J. McHenry, personal communication, March, 1986).

Table 29. Split-Half and Test-Retest Reliability of the PAPB Computer Battery Tests

Test	Split-Half Reliability	Test-Retest Reliability
Tracking 1 (one-handed tracking)	.98	.74
Tracking 2	.98	.85
Target Shoot (distance measure)	.74	.37
Target Shoot (time to fire)	.85	.58
Cannon Shoot (time discrepancy)	.65	.52

Selection of Interest and Temperament Measures

Based on recommendations from Dr. Leaetta Hough (personal communication, January, 1987) we decided to use seven scales from the ABLE and AVOICE. Three scales were selected from the AVOICE: Combat, Rugged Individualism, and Fire Arms Enthusiast. These

are the three scales determined to related to the "Combat" construct. Four scales were selected from the ABLE: Dominance, reflecting the construct of ascendancy; Traditional Values and Nondelinquency, two scales reflecting dependability; and Non-Random Response, a response validity scale.

Selection of ASVAB Composites

We selected two ASVAB composites to use in the ET/NT/ANCOC profile comparison, Combat Operations (CO) and General Technical (GT). CO has proven predictive of performance in training and in units on first assignment (Black, 1980; Campbell & Black, 1982). GT was selected based on the recommendation of Dr. Scott Graham (personal communication, August, 1988) as a measure frequently investigated in the prediction of Armor performance.

Thus, in sum we used eleven variables in the ET/NT/ANCOC comparison: two ASVAB composites, CO and GT; two measures from the PAPB psychomotor battery, Factor 1 and Factor 2; and seven scales from the ABLE and AVOICE, Combat, Rugged Individualism, Fire Arms Enthusiast, Dominance, Traditional Values, Nondelinquency, and Non-Random Response. It might be noted that the ASVAB composite CO and Factor 1 from the PAPB are the cognitive and psychomotor matching variables used in our ET program evaluation.

Research Participants

The profile comparison required defining a sample of ETs, NTs, and ANCOCs. All research participants were U.S. Army enlisted personnel undergoing Armor training at Fort Knox, Kentucky. The ET participants were the 83 ETs selected from A, B, C companies across the five cycles of OSUT between May and December 1986. These are the same ET soldiers used in our ET program evaluation described previously. Some 83 NCOs undergoing M1 Advanced NCO Courses (M1 ANCOCs) at Fort Knox during the months of June and October 1986 served as our ANCOC participants.

The group of NTs used in the profile comparison could not be the same as those matched with ETs for the purposes of describing training program criterion performance differences. Instead, this sample of NTs needed to be representative of the aptitude, interest, and temperament distribution of all NTs. Thus 41 NT cohorts were randomly sampled from the same five companies from which our ET participants were drawn. Forty-one NTs were selected to parallel the sample size for ANCOCs for whom we had complete data, as detailed below.

Collecting PAPB and ASVAB Data

As indicated previously, the profile comparison approach involved obtaining ASVAB scores and PAPB data from a sample of ETs, NTs, and ANCOCs. The PAPB data for the ET and NT soldiers were collected during Week 8 of OSUT for each cycle of soldiers, as described previously in the discussion of the ET program evaluation. The ASVAB scores were likewise obtained for ET and NT soldiers during the first weeks of OSUT while their records were at Fort Knox. We obtained ASVAB scores for all 83 ET and 41 NT participants. We collected PAPB scores for all 83 ET participants and for 40 of the 41 NT participants.

PAPB data were collected from ANCOCs in a similar manner to that described for the ET/NT data collection. These soldiers, while at Fort Knox undergoing M1 ANCOC training, were tested following the standard PAPB administration procedures. However, their ASVAB records were not on post at the time of testing. ANCOC ASVAB scores were obtained through computerized central record keeping. From the 83 ANCOCs tested, we obtained valid PAPB data on 72 soldiers and obtained ASVAB scores for 41 of these 72 soldiers. We were not successful at locating ASVAB scores for 31 of the ANCOCs for whom we had PAPB scores. Thus, we have both ASVAB and PAPB data on only 41 of our ANCOC soldiers.

ET/NT/ANCOC Profile Comparisons

Our analysis of the profile similarity among these three groups of soldiers began by computing the mean and standard deviation for each group on the two ASVAB scales, the two PAPB psychomotor factors, and each of the seven recommended PAPB ABLE/AVOICE scales. The results of these computations are displayed in Table 30. All measures are reported in standard score form based on the entire population. ASVAB GT and CO are c-scores, the remaining measures are z scores.

Table 30. ET/NT/ANCOC Aptitude/Interest/Temperament Scale Means & SDs.

Profile Measures	Soldier Classification		
	ET	NT	ANCOC
GT			
Mean	111	107	104
Standard Deviation	11	9	17
Valid N	84	41	41
CO			
Mean	115	110	109
Standard Deviation	11	8	16
Valid N	85	41	41
FACTOR1			
Mean	.715	.480	-.239
Standard Deviation	.751	.888	.968
Valid N	85	40	67
FACTOR2			
Mean	.383	.509	-.007
Standard Deviation	.674	.781	.751
Valid N	85	40	67
NON-DELINQUENCY			
Mean	.29	.42	-.37
Standard Deviation	1.58	.98	1.75
Valid N	84	41	72
TRADITIONAL VALUES			
Mean	.51	.61	.27
Standard Deviation	.93	.97	.83
Valid N	84	39	71
DOMINANCE			
Mean	.51	.48	.60
Standard Deviation	1.07	1.00	.93
Valid N	84	41	72

Table 30. ET/NT/ANCOC Aptitude/Interest/Temperament Scale Means & SDs (continued)

Profile Measures	Soldier Classification		
	ET	NT	ANCOC
COMBAT			
Mean	1.38	1.35	.88
Standard Deviation	.66	.79	.96
Valid N	84	41	72
RUGGED INDIVIDUALISM			
Mean	.49	.40	.21
Standard Deviation	.78	.86	.91
Valid N	84	41	72
FIRE ARMS ENTHUSIAST			
Mean	.44	.42	.45
Standard Deviation	.82	.93	1.10
Valid N	84	41	72
VALIDITY			
Mean	.23	-.07	-.48
Standard Deviation	.94	1.59	1.70
Valid N	84	41	72

It is apparent that ETs are indeed a select group. On all the aptitude measures they score substantially above the mean. It is also clear that ANCOCs are not a particularly select group based on these aptitude measures. ANCOCs actually score below the mean on the psychomotor measures, though it must be remembered that these measures were normed on younger, first tour soldiers. NTs appear to fall somewhere in between ETs and ANCOCs on the aptitude measures.

ETs also score high on a number of the ABLE/AVOICE scales. Most notable is their score on the Combat scale of over one standard deviation above the mean. However, as on the aptitude scales, here too, it appears that there is greater similarity between ETs and NTs than there is between ETs and ANCOCs.

A visual presentation of the relationship among the three profiles is provided in Figure 3. Each measure has been converted to a T score based on its mean and standard deviation for the three groups combined. Neither the shape nor the level of the profiles appears to support the proposition that ETs and ANCOCs' profiles are more similar than NT and ANCOC profiles. In fact, there appears to be greater similarity between the latter two groups.

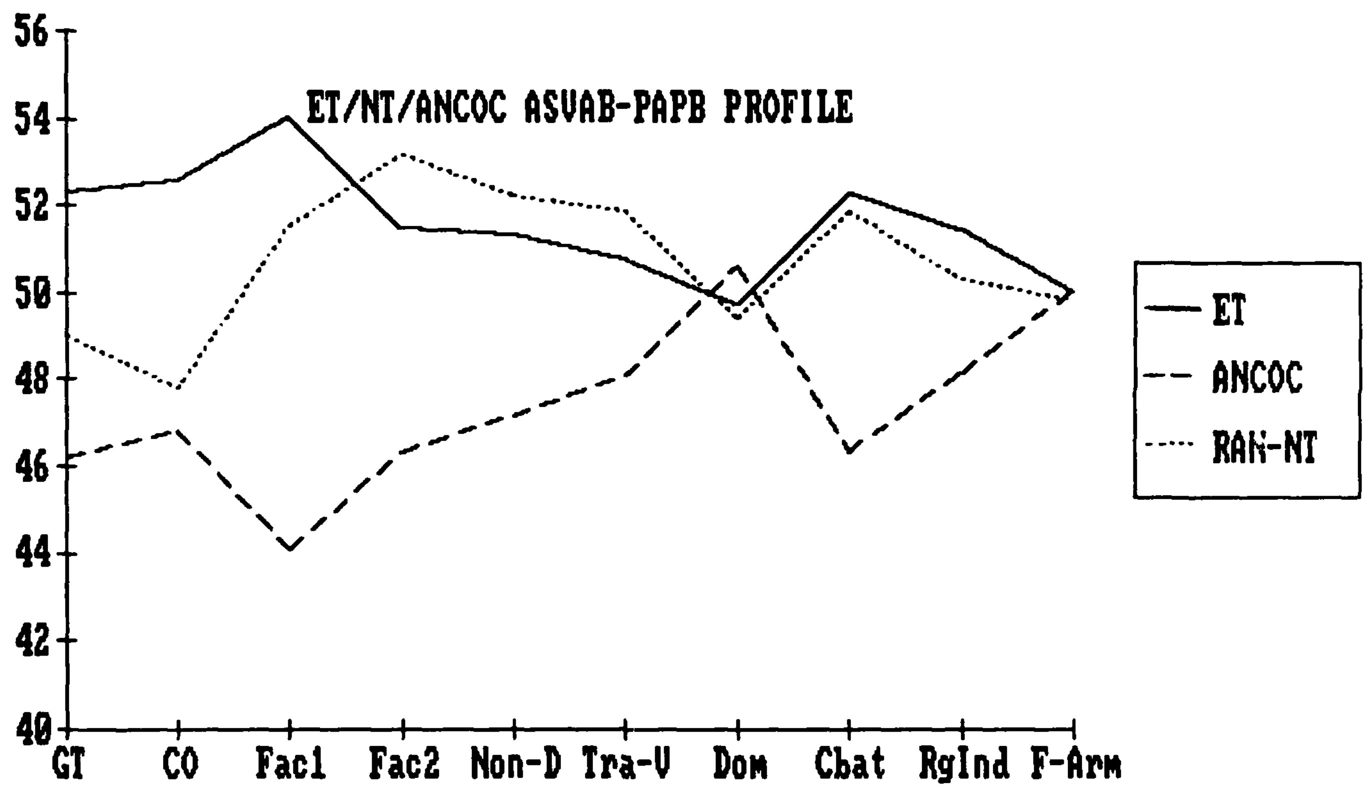

Figure 3. ET, NT, and ANCOC profile comparison.

Multiple discriminant function analyses were performed in order to further evaluate the degree of similarity among the profiles. These analyses were performed in two ways. First, we simply evaluated the extent to which the eleven profile measures can be used to correctly predict membership in one of the three groups. Then we developed a model designed to differentiate between NTs and ANCOCs. The resulting model was subsequently used to classify ETs. The results of both analyses are reported below.

When attempting to differentiate among the three groups using the eleven predictors identified above, two discriminant functions were extracted, but only the first function is statistically significant ($p<.001$). The canonical correlation between the first discriminant function and the group designation is .49. Table 31 displays the correlation between each predictor and the first discriminant function score. As can be seen, Factor 1, a psychomotor composite from the PAPB best differentiates among the groups, followed by the Combat scale from the AVOICE and the GT scale from the ASVAB.

Table 31. Correlations between Predictors & Discriminant Function Score: Three Group Model

Predictor	Correlation
FACTOR1	.77634*
C	.43830*
GT	.40257*
NR	.34171*
R	.18032*
CO	.38102
FACTOR2	.26979
ND	.21278
T	.16756
F	-.00433
D	.01330

Employing the above model to predict group membership produced the results displayed in Table 32. Approximately 52% of the soldiers were correctly assigned to their actual group based on their profile scores. Though this clearly is an improvement over the chance rate of 33%, classification is by no means accurate.

Table 32. Multiple Discriminant Function Analysis Classification Table: Three Group Model

Actual Group Membership	No. of Cases	Predicted Group Membership ET	NT	ANCOC
Group ET	83	45 54.2%	25 30.1%	13 15.7%
Group NT	38	13 34.2%	15 39.5%	10 26.3%
Group ANCOC	39	10 25.6%	6 15.4%	23 59.0%

Percent of "grouped" cases correctly classified: 51.88%

It is informative to examine the nature of the classification errors. Forty-six percent of ETs were misclassified. Thirty percent of the misclassifications resulted from classifying the ETs as NTs. Only 16% of the ETs were "confused" with ANCOCs. These results confirm the notion that ETs' profiles look more like NT profiles than ANCOC profiles. However, when we examine the misclassification errors for ANCOCs, we see greater confusion of ANCOCs with ETs (26%), than with NTs (15%). This suggests ANCOCs and ETs are more similar. To help resolve these seemingly inconsistent results, we examined the classification of ETs based on a model designed to optimally differentiate between NTs and ANCOCs.

A significant discriminant function did result from regressing just the NT/ANCOC group designation on the predictor profiles ($p<.025$). Table 33 displays the correlations between the predictors and the discriminant function score. The relative usefulness of the predictors in distinguishing between NTs and ANCOCs parallels the previous three group analysis.

Table 33. Correlations between Predictors & Discriminant Function Score: Two Group Model

Predictor	Correlation
FACTOR1	.61840
C	.43565
GT	.41664
NR	.41279
R	.26172
CO	.18368
FACTOR2	.17705
ND	.14578
T	.07555
F	.02486
D	.02083

Contrasted with a chance rate of 50%, Table 34 reveals 71% of the NTs and ANCOCs were correctly classified. The important result from this analysis, however, is the classification of ETs. Here 75% of the ETs are classified as NTs, only 25% are assigned to the ANCOC group. This offers considerable support for the notion that ET profiles are more similar to those of NTs than they are similar to ANCOCs'.

Table 34. Multiple Discriminant Function Analysis Classification Table: Two Group Model

Actual Group Membership	No. of Cases	Predicted Group Mem NT	ANCOC
Group NT	38	27 71.1%	11 28.9%
Group ANCOC	39	11 28.2%	28 71.8%
Ungrouped Cases ET	83	62 74.7%	21 25.3%

Note: Percent of "grouped" cases correctly classified: 71.43%

One caution must be noted. The above analyses involve a relatively large number of predictors (i.e., eleven) in relation to the relatively small number of soldiers in each group. The likely consequence of this is to overstate the value of the profile variables for distinguishing among the above groups. However, since even these results do not provide support for the proposition that ET profiles more closely resemble ANCOC profiles than do NT profiles, it is doubtful that increasing the sample size would alter this conclusion.

PREDICTION OF OSUT PERFORMANCE

A final, post hoc issue addressed in this research effort is the extent to which various OSUT performance composites are predictable from selected ASVAB and PAPB measures. A primary concern is the predictability of UCOFT performance. Accordingly we address this issue first. Then we examine the predictability of performance composites derived from TCGST, Military Stakes, and the Project A rating scales.

Research Participants

Data from all soldiers for whom we had ASVAB, PAPB, and reasonably complete criterion performance measures were used in this phase of the research. These 124 soldiers included all 83 ETs. It also included 41 NTs from our matched group who were also selected as part of the group of randomly sampled NTs. These were the only NTs for whom we had scored AVOICE and ABLE protocols.

Criterion Performance Composites

Previously we identified and discussed the rationale for a large number of performance measures gathered to evaluate ET/NT performance during OSUT. In particular, we identified multiple measures of performance on the UCOFT, TCGST, Military Stakes, NT Paper and Pencil Test, and Project A Army-Wide Rating Scales. Conducting separate regression analyses for each of the component measures within each of the above domains would make little sense.

Analyses of the intercorrelations among the component measures within a domain indicated that it would be reasonable to combine the within domain measures. Accordingly, a composite for each domain was formed by standardizing each component measure within a domain, then simply summing the standard scores. This procedure was strictly followed to develop the TCGST composite (TCGST) and the Project A Army-Wide Rating Scales composite (RATING). The Military Stakes composite (MILSTAKE) was the sum of seven standardized components, the six Military Stakes composites previously described and the NT Paper and Pencil Test (NTPP). The single NTPP score was not analyzed separately, since like the Military Stakes, it was designed to sample the NT performance domain. However, to avoid having the six Military Stakes measures mask the contribution of the NTPP, in forming the MILSTAKE composite, we weighted the NTPP standard score by three. A UCOFT performance composite (COFT) was formed by equally weighting the Accuracy, Latency, and System Management scores. Given the particular interest in UCOFT performance, we also examine the prediction of Accuracy, Latency, and System Management individually.

For a variety of reasons, some soldiers were missing an occasional score on one or two of the components within one or more of the above domains. In order to avoid excessive subject attrition, where a soldier had two or fewer missing scores within any of the above domains, the item mean was substituted for the missing data point before forming the domain composite. The impact of this and the restricted predictor and criterion variance resulting from non-random soldier selection - 83 of the 123 soldiers were ETs - will generally result in the under estimation of true validity.

Identification of Trial Predictors

For the present analyses, there are available a large number of potential predictors from the ASVAB and PAPB, but only a limited number of soldiers on whom predictor and criterion data are available. Consequently, generalizable results can be obtained only by pre-selecting a subset of predictors for consideration in the regression analyses.

With this in mind, the predictors we used were the GT and CO scales from the ASVAB, the three factor scores from the computer-based PAPB psychomotor tests (hereafter identified as Factor 1, Factor 2, & Factor 3), and a Combat composite and a Dependability composite derived from the AVOICE and the ABLE, respectively.

The two ASVAB scales were selected based on research cited earlier suggesting these measures are predictive of gunnery performance. The three psychomotor factors were selected for the hypothesized construct validity of these measures for gunnery tasks. The Combat Composite is the sum of the scores on the Combat, Rugged Individualism, and Fire Arms Enthusiast AVOICE subscales. The Dependability Composite is the sum of the Traditional Values scale (T) and the Non-Delinquency scale (ND) in the ABLE. These two composites were chosen because they demonstrated validity predicting core technical proficiency of 19Es in a preliminary examination of their validity for Project A (Campbell, 1986).

The resulting matrix of intercorrelations among all available ASVAB, PAPB, and UCOFT scores is provided in the Appendix.

Prediction of UCOFT Performance

This section addresses two issues regarding the prediction of UCOFT performance. One question concerns the relative contributions of the selected ASVAB versus the PAPB measures to the explanation of UCOFT performance variance. The second issue is the moderating impact of engagement degradation on the above validity coefficients.

To answer the first question, we performed four separate stepwise regressions, one for each of the UCOFT criterion performance composites developed for the evaluation of ET/NT UCOFT performance differences. Thus, the dependent variables are Accuracy, Latency, System Management, and the equally weighted composite of these three measures entitled COFT. The predictors available for inclusion in the stepwise analysis were the seven ASVAB/PAPB measures identified above. These data were available for 124 soldiers.

Table 35 shows that for Accuracy, when using a partial F-test to determine at what point additional predictors fail to significantly contribute to the amount of explained variance, only the PAPB psychomotor tracking factor (Factor 1) and the AVOICE Combat Composite entered the equation. Together these two measures produced a multiple R of only .23. Correcting this validity for the marginal reliability of the Accuracy measure (rel = .58), increases the validity to R = .30.

Table 35. Regression of UCOFT Accuracy Score on ASVAB/PAPB

Accuracy						
Variable	Step	MultR	Rsq	F(Eqn)	SigF	BetaIn
FACTOR1	1	.2323	.0540	6.961	.009	.2323
COMBAT	2	.3043	.0926	6.176	.003	.1966

------------------ Variables in the Equation -----------------

Variable	B	SE B	Beta	T	SigT
FACTOR1	3.24670	1.22932	.22874	2.641	.0094
COMBAT	.20624	.09085	.19663	2.270	.0250
(Constant)	37.73719	4.69346		8.040	.0000

The Latency composite is predicted much better. Table 36 shows a multiple R= .50 results when a four predictor model is defined. Corrected for the unreliability of the Latency measure (rel = .86), the validity coefficient increases to R = .60. The majority of the explained variance is attributable to the tracking factor from the PAPB. Notably, both temperament composites from the PAPB enter the equation. Here the CO scale from the ASVAB also explains a small amount of variance in addition to that accounted for by the PAPB.

Table 36. Regression of UCOFT Latency Score on ASVAB/PAPB

Latency						
Variable	Step	MultR	Rsq	F(Eqn)	SigF	BetaIn
FACTOR1	1	.3528	.1245	17.344	.000	.3528
COMBAT	2	.4276	.1828	13.534	.000	.2416
DEPEND	3	.4735	.2242	11.559	.000	-.2058
CO	4	.5052	.2553	10.197	.000	.1907

------------------ Variables in the Equation ------------------

Variable	B	SE B	Beta	T	Sig T
FACTOR1	4.28694	1.18517	.30680	3.617	.0004
COMBAT	.29474	.08283	.28544	3.558	.0005
DEPEND	-.23393	.08354	-.22548	-2.800	.0060
CO	.18734	.08407	.19074	2.228	.0277
(Constant)	22.85173	10.90605		2.095	.0383

System Management performance is also predicted well by the two PAPB psychomotor factor scores. The multiple R in Table 37, corrected for the unreliability of the System Management composite, is R = .73. Somewhat surprising is the failure of CO or GT to enter the model given the seemingly greater cognitive nature of this composite.

Table 37. Regression of UCOFT System Management Score on ASVAB/PAPB

System Management						
Variable	Step	MultR	Rsq	F(Eqn)	SigF	BetaIn
FACTOR1	1	.4625	.2139	33.205	.000	.4625
FACTOR2	2	.5085	.2586	21.104	.000	.2753

------------------ Variables in the Equation ------------------

Variable	B	SE B	Beta	T	Sig T
FACTOR1	4.17955	1.48920	.28614	2.807	.0058
FACTOR2	4.56951	1.69221	.27531	2.700	.0079
(Constant)	45.15787	1.16356		38.810	.0000

The results of the regression of the UCOFT overall composite performance measure on the seven predictors are reported in Table 38. Again a considerable proportion of criterion variance is explained. Once again the PAPB tracking factor is the best predictor. It is clear that, while the ASVAB CO scale does correlate with UCOFT performance, the combination of the PAPB psychomotor and temperament scales accounts for most of the explained criterion variance.

Table 38. Regression of UCOFT Composite Score on ASVAB/PAPB

COFT						
Variable	Step	MultR	Rsq	F(Eqn)	SigF	BetaIn
FACTOR1	1	.4119	.1697	24.929	.000	.4119
COMBAT	2	.4588	.2105	16.131	.000	.2021
FACTOR3	3	.4902	.2403	12.654	.000	.1832
CO	4	.5158	.2661	10.785	.000	.1726

------------------ Variables in the Equation ------------------

Variable	B	SE B	Beta	T	Sig T
FACTOR1	4.18344	1.24849	.29327	3.351	.0011
COMBAT	.26382	.08454	.25026	3.121	.0023
FACTOR3	2.17171	.99281	.18224	2.187	.0307
CO	.17303	.08470	.17256	2.043	.0433
(Constant)	13.75869	10.75578		1.279	.2033

Prediction of Normal vs. Degraded UCOFT Performance

Results cited earlier suggest the differences between ET and NT UCOFT performance are greater under degraded than under normal UCOFT engagements. This gives rise to the question of whether or not different abilities are involved in UCOFT performance under these two conditions. While the sample size limitations of the present data base preclude a definitive answer to this question, we can at least take a preliminary look at this issue.

One way of approaching this question is to examine the relationship between the predicted scores derived from the regression of normal UCOFT performance on the seven predictors identified earlier with the predicted scores derived from a parallel analyses on degraded UCOFT engagements. If the correlation between these vectors approaches unity then there is little reason to believe a different mix of abilities is involved in gunnery performance under these two conditions. On the other hand, if these vectors are not highly correlated, this may suggest different abilities underlie performance.

We began by performing eight separate regressions. The dependent variables were accuracy, latency, system management, and the COFT composite on both normal and on degraded exercises. To ensure the same model was invoked each time, all seven predictors identified earlier were entered into the prediction model. The resulting multiple correlations and omnibus p-values are displayed in Table 39. With the exception of accuracy during normal exercises, substantial prediction of all UCOFT performance measures under both modes was achieved. It is also apparent that there is little difference in the relative sizes of the multiple correlations across modes. However, each equation was optimized to predict a different criterion.

Table 39. Prediction of Normal and Degraded UCOFT Exercise Performance: Multiple Correlations

	Normal		Degraded	
UCOFT Criterion	R	p<	R	p<
Accuracy	.289	.168	.404	.004
Latency	.529	.000	.496	.000
System Management	.446	.000	.513	.000
COFT Composite	.497	.000	.512	.000

The matrix of intercorrelations among the predicted scores derived from the eight regression equations is shown in Table 40. The underlined coefficients denote the correlations between the vectors of like criterion measures. The high correlations between the predicted "normal" criterion and the corresponding predicted "degraded" criterion suggests the predicted rank order of soldiers based on their performance under the two modes is similar.

Table 40. Intercorrelations Among Predicted Normal and Predicted Degraded Mode UCOFT Performance Measures

Normal Exercises	Degraded Exercises			
	Accuracy	Latency	Sys. Mgmt.	COFT
Accuracy	.8225	.8789	.8051	.9352
Latency	.6595	.8984	.9505	.9544
Sys. Mgmt.	.6258	.7891	.9183	.8875
COFT	.7146	.8895	.9408	.9636

Yet another approach to examining the similarities and differences among the normal and degraded mode equations is to view the problem as a variant of cross-validation. Instead of evaluating the stability of the regression model across samples, here we are interested in the shrinkage in explained variance across criteria. That is, how well does an equation developed on a normal mode criterion predict the same criterion under degraded conditions? The parallel question can be asked with regard to the prediction of normal engagements from an equation developed on degraded exercises. In Table 41 we report the results of four double cross validation analyses, one for each of the *UCOFT* criteria.

Table 41. Validities & Cross-Validities *for* Prediction of UCOFT Performance Under Normal and Degraded Conditions

Predictor Composite	UCOFT Criterion Composite NAC	DAC	NLAT	DLAT	NSYS	DSYS	NCOFT	DCOFT
NAC	.2896	.3325	.4498	.4364	.3562	.4134	.4496	.4792
DAC	.2382	.4042	.3490	.3960	.2793	.2577	.3558	.4251
NLAT	.2461	.2666	.5292	.4461	.4262	.4880	.4912	.4890
DLAT	.2545	.3224	.4755	.4965	.3521	.3958	.4429	.4924
NSYS	.2312	.2530	.5055	.3918	.4462	.4715	.4836	.4547
DSYS	.2332	.2029	.5030	.3828	.4098	.5134	.4684	.4493
NCOFT	.2615	.2889	.5221	.4417	.4334	.4830	.4979	.4937
DCOFT	.2708	.3354	.5051	.4772	.3960	.4502	.4798	. 5124

The four coefficients within each box are the validities and the cross-validities for each criterion. Thus, the equation developed to predict accuracy in normal engagements (.2896), predicts accuracy under degraded conditions somewhat better (.3325). The degraded accuracy equation however, predicts degraded accuracy (.4042) better than it predicts accuracy under normal engagements (.2382). The pattern of coefficients within each box reveals only a trivial loss in predictive efficiency when an equation optimized for one criterion is applied to the other criterion. In general, the relationship among the relevant coefficients within each block offer little evidence to suggest different equations, and thus a different mix of abilities underlies successful performance in normal versus degraded engagements.

Prediction of TCGST, Military Stakes, Project A Ratings and Overall Composites

Tables 42 through 45 display the results of the four stepwise regressions of the TCGST, Military Stakes, Project A Rating, and Overall composites respectively on the seven ASVAB/PAPB predictors. With the exception of the TCGST composite, substantial variance was explained in these remaining OSUT performance measures. Though CO was a less powerful predictor of UCOFT performance, here the CO composite from the ASVAB enters the prediction model for all four criteria below and it accounts for the majority of variance in three of the four equations. Thus the robustness of the CO scale for predicting many aspects of OSUT performance of tank crewman is again demonstrated.

The multiple correlation of .29 shown in Table 42 for the prediction of TCGST performance is low. In all likelihood, this is in large measure a result of a rather unreliable criterion. In addition, as mentioned earlier, administrative irregularities associated with the gathering of the components of the TCGST composite undermine our confidence in the integrity of this measure.

Table 42. Regression of TCGST Composite Score on ASVAB/PAPB

TCGST Composite						
Variable	Step	MultR	Rsq	F(Eqn)	SigF	BetaIn
CO	1	.2986	.0892	11.943	.001	.2986

------------------ Variables in the Equation ------------------

Variable	B	SE B	Beta	T	Sig
CO	.29158	.08437	.29860	3.456	.0008
(Constant)	17.69021	9.64598		1.834	.0691

Table 43 shows the substantial multiple correlation between the predictor composite and the Military Stakes composite ($R = .549$) is attributable entirely to the two ASVAB measures GT and CO. None of the PAPB psychomotor or temperament/interest scales entered the stepwise analyses. Given the substantial correlation between these two ASVAB scales, little can be said about their contribution relative to each other.

Table 43. Regression of Military Stakes Composite Score on ASVAB/PAPB

Military Stakes Composite						
Variable	Step	MultR	Rsq	F(Eqn)	SigF	BetaIn
GT	1	.5492	.3016	52.245	.000	.5492
CO	2	.5750	.3307	29.641	.000	.2418

Variables in the Equation					
Variable	B	SE B	Beta	T	Sig T
GT	.34787	.09752	.37772	3.567	.0005
CO	.22442	.09826	.24183	2.284	.0241
(Constant)	-13.94790	8.42607		-1.655	.1005

As shown in Table 44, prediction of the composite of supervisor and peer overall performance ratings was achieved by combining the CO scale from the ASVAB with the Dependability factor from the PAPB ABLE. It is noteworthy that Dependability, comprised of the ABLE Non-Delinquency and Conscientiousness scales, explains substantial rating variance beyond that accounted by the cognitive ability predictor CO. It appears that peers and supervisors were indeed responding to both cognitive and temperament aspects of job performance when completing their overall soldiering performance ratings.

Table 44. Regression of Project A Ratings Composite on ASVAB/PAPB

Project A Ratings Composite						
Variable	Step	MultR	Rsq	F(Eqn)	SigF	BetaIn
CO	1	.3569	.1274	17.807	.000	.3569
DEPEND	2	.4572	.2090	15.985	.000	.2882

Variables in the Equation					
Variable	B	SE B	Beta	T	Sig T
CO	.29958	.07648	.31943	3.917	.0001
DEPEND	.28544	.08078	.28815	3.534	.0006
(Constant)	3.47309	9.14156		.380	.7047

The equally weighted composite formed from the UCOFT, TCGST, Military Stakes including the NT Paper and Pencil Test, and Project A peer and supervisor ratings was predicted quite well by CO, as shown in Table 45. Factor 2 from the PAPB psychomotor battery accounted for some additional variance. The very substantial predictive power of the ASVAB CO score is not surprising in view of the cognitive demands placed upon soldiers during OSUT. As is apparent however by comparing Tables 35 through 38 with Tables 41 through 45, the greater the emphasis specifically upon gunnery performance, the greater is the importance of the PAPB psychomotor and temperament/interest factors.

Table 45. Regression of OSUT Performance Composite on ASVAB/PAPB

OSUT Performance Composite						
Variable	Step	MultR	Rsq	F(Eqn)	SigF	BetaIn
CO	1	.5128	.2630	43.177	.000	.5128
FACTOR2	2	.5577	.3110	27.085	.000	. 2273

------------------ Variables in the Equation ------------------

Variable	B	SE B	Beta	T	Sig T
CO	.44814	.07788	.45231	5.754	.0000
FACTOR2	3.64288	1.25957	.22735	2.892	.0045
(Constant)	-1.43810	8.77850		-.164	.8701

DISCUSSION

The present research effort had three major objectives: a) compare ET/NT OSUT performance with particular emphasis on gunnery proficiency measured in the UCOFT , b) examine the aptitude/interest/temperament profile similarities among ETs, NTs, and ANCOCs, and c) investigate the validity of selected ASVAB and PAPB scales for the prediction of OSUT performance. The design, analyses, and results obtained in addressing each of these objectives were presented in the previous sections. Here we briefly review and evaluate the results pertaining to each objective.

ET/NT OSUT Performance: The Impact of the Excellence Track

The multiple measures collected on performance during OSUT convincingly demonstrate the impact of the Excellence Track. On a wide variety of indicators we found soldiers in the Excellence Track become both more knowledgeable and more skillful than their cohorts who are not exposed to the Excellence Track POI. Because the cohorts in this investigation were comparable to their ET counterparts in cognitive and psychomotor ability, we are confident that the performance gains associated with ETs can be attributed directly to the structure and content of the Excellence Track program.

The reader will recall that performance measures were developed to sample different aspects of OSUT content. Thus the UCOFT and TCGST measures disproportionately sampled Excellence Track content, the Military Stakes and NT Paper and Pencil Test targeted the content of the Normal Track, and the Project A Performance Ratings sampled the domain common to both tracks. We believe this measurement strategy was necessary because of the possibility that the Excellence Track might produce performance gains in areas emphasized in the Excellence Track POI at the expense of knowledge and skill development in areas receiving more attention in the Normal Track. This appeared to be a reasonable outcome since all the time devoted to the Excellence Track POI is captured by requiring ETs to master the Normal Track content in far less time.

However, a particularly encouraging finding is the absence of an ET performance decrement in those content areas for which much less training time is allocated relative to the Normal Track. Despite the compressed time frame ETs had available to master this content, they performed marginally better on the Military Stakes and substantially better on the NT Paper and Pencil Test than did their cognitively matched NT cohorts.

The performance ratings reveal very substantial ET superiority. Since the rating dimensions targeted by these scales refer to Army-wide general soldiering proficiencies, they do not inherently favor either group. It is not clear what aspects of the Excellence Track program could be expected to produce these ET/NT differences. Three explanations seem plausible. It may be simple rater bias, since the peer and supervisor raters knew who were ETs and who were NTs. Alternatively, the attributes reflected in these ratings may have been present at the time of ET selection, and our matching procedure on cognitive and psychomotor ability failed to equate the groups on these dimensions. Finally, it is reasonable to speculate that selection for and participation in the Excellence Track creates an esprit de corps that fosters the development of the qualities measured by these scales. Further investigation is required if we wish to be in a position to eliminate one or more of these competing explanations.

In view of the above findings, it is not surprising that ETs excelled in those areas where they received more focused training. They clearly outperformed NTs on both the UCOFT and the TCGST. On the UCOFT, although they did not respond more quickly, they were more accurate and made fewer system management errors. This difference resulted despite the fact that our ETs enjoyed no advantage over the NT matches in actual hands-on UCOFT experience.

What may be particularly significant about the ET/NT UCOFT performance differences is where the superior gunnery performance occurs. The analyses of normal exercise versus degraded exercise UCOFT performance revealed that the ET performance superiority is manifest under degraded rather than normal conditions. It has been argued that the degraded exercises are far more realistic, that is, it is under these conditions battles are actually fought. If this is so, then the enhanced performance of Excellence Track soldiers in this mode is an all the more compelling endorsement of the Excellence Track program.

The large differences between ETs and NTs on the TCGST are reassuring, but not surprising. Among our criterion measures, the TCGST most favors ETs since they, not NTs, receive training on the content measured by the TCGST. What this measure does show is that the Excellence Track training does result in an added domain of knowledge and skill acquisition.

In summary, the Excellence Track appears highly successful in achieving its intended purpose. A broad range of additional knowledges and skills over those acquired in the Normal Track are clearly evident. This performance gain is acquired without sacrificing performance in the NT domain. In fact, ET performance on the NT domain actually appears to be enhanced as well through participation in the Excellence Track.

A comment or two is also in order regarding the usefulness of the UCOFT as a device for gathering information about gunnery performance. Two issues deserve mention: reliability and TC contamination. Careful attention was given to selecting exercises that would produce reasonable performance variance. In addition, dispersion rounds were not counted. These steps were taken because prior research identified these actions as prerequisites to obtaining reliable UCOFT measures. Nevertheless, the resulting test-retest reliability coefficients were still lower than is desirable for performance measurement. The reliability analysis indicates that, at least for inexperienced gunners, rather lengthy testing sessions are needed in order to get reliable measures of UCOFT gunnery performance. The reader is reminded however, that regarding the description of ET/NT gunnery performance differences, the attenuated reliability has only served to understate the magnitude of the observed UCOFT differences. That is, in all probability, ETs' gunnery performance gains over NTs are actually larger than reported here.

The other lingering difficulty with the UCOFT as a device for measuring gunner performance is the contaminating role of the tank commander. Despite our extensive efforts to standardize the performance of the TC through both training and performance monitoring, seemingly very similar TC performance still had a marked impact on all three gunner's performance composites. While the design of the present investigation successfully isolated this effect, the problem remains for more routine future attempts to measure gunner performance on the UCOFT. However, the recently developed Institutional Conduct of Fire Trainer (ICOFT) has corrected this contaminant in the evaluation of gunnery performance by automating the role of TC.

Comparison of ET/NT/ANCOC Aptitude/Temperament/Interest Profiles

Our second research objective, investigating the profile similarities between Excellence Track soldiers and NCOs produced less promising results. The hope was that the future early identification and selection of candidates for the Excellence Track could be facilitated by this effort. If ETs' ASVAB/PAPB profiles are similar to NCOs' profiles and at the same time dissimilar from NTs' profiles, than cadre could look to these measures as an aid in the identification of promising prospects both for the Excellence Track and as soldiers more likely to eventually become NCOs. The data examined in the present investigation do not support the notion that ETs and NCOs have more similar profiles than NTs and NCOs. On the contrary, Normal Track soldiers profiles are more similar to NCOs than are Excellence Track soldier's profiles.

Somewhat disturbing is the rather unflattering picture the data provide of NCOs, at least the sample of 41 NCOs participating in this investigation. On both the cognitive and the psychomotor measures, the NCOs performed less well than the NTs and substantially less well than the ETs. While the pattern is somewhat less interpretable for the interest and temperament measures, here too there tends to be greater similarity between ETs and NTs than between either of these groups and NCOs. For example, on the Combat and Fire Arms Enthusiast Scales which proved predictive of several OSUT performance measures, NCOs score substantially lower than either ETs or NTs. In essence, there is a hierarchy of cognitive and psychomotor ability with ETs at the top and NCOs at the bottom. On the relevant ABLE/AVOICE scales, particularly the two just mentioned, the profiles break out the same way. Although the latter scales do reflect neither correct nor incorrect answers, it is reasonable to expect enthusiastic Combat Arms soldiers to receive relatively high scores on these measures.

How do we account for these results? One possibility is that at least our PAPB measures are distorted. Many of the ANCOCs appearing for PAPB testing for this project reportedly expressed considerable dissatisfaction over their participation. They clearly did not want to be tested. If as a result they malingered, their aptitude scores would be depressed and quite possibly their interest/temperament scores might be distorted as well. Some support for this hypothesis is gleaned from the validity scale in the ABLE. Here, the ANCOCs scored a half standard deviation below the mean. Thus we clearly have reason to suspect these scores. However, the ANCOCs ASVAB scores were also lowest of the three groups and these scores were culled from their military records dating back to their initial entry into the Army. Hence, on balance, we believe we at least have an undistorted picture of the cognitive and psychomotor abilities of this group of NCOs. The temperament/interest profile for this group is suspect.

Another possibility is the group of ANCOCs we happened to select, or the subset of this group for whom we successfully recovered ASVAB scores are unrepresentative of NCOs in this MOS. We presently do not have any information available that permits us to explore the reasonableness of this hypothesis.

A final explanation is simply that the soldiers that remain in the Army and progress to NCOs in this area just are not a particularly gifted group. Perhaps, at the time these NCOs were making career decisions, better alternatives were available to their smarter cohorts.

Regardless of the explanation for the NCO profiles obtained in this investigation, the present data do not indicate that participation in the Excellence Track should turn on the degree of aptitude/temperament/interest profile similarity between OSUT soldiers and NCOs. As we indicate in the next section, better data are available to facilitate the ET selection process.

Prediction of OSUT Performance

Selected scales from the ASVAB and the PAPB proved to be very effective predictors of a broad range of OSUT performance measures. Especially useful were the CO composite from the ASVAB and the tracking composite, Factor 1, from the computer administered psychomotor component of the PAPB.

Not surprisingly, gunnery performance as measured on the UCOFT was predicted well by Factor 1. Bivariate correlations between Factor 1 and the four UCOFT performance measures ranged from $r = .23$ to $r = .46$. Given the attenuated reliability of these UCOFT performance measures and the predictor range restriction resulting from the ET selection and NT matching processes, these validities are quite impressive. Their true validity is substantially higher. Additional predictors entering the stepwise procedure were also generally from the PAPB. The ABLE Combat and Dependability scales each accounted for significant increments in explained variance. It is apparent that the PAPB has accomplished its intended purpose of extending the prediction of soldier performance beyond that achieved with the ASVAB.

As in prior studies, the effectiveness of CO as a predictor of gunnery performance is unclear. Though CO does enter the model for prediction of the UCOFT COFT composite and the Latency composite, in each case it is the last of four predictors to enter the model and the increment in explained variance is small. To be sure one must be cautious when interpreting the order of entry of correlated predictors into a regression model, but inspection of the magnitude of the bivariate correlations reveals that Factor 1 from the PAPB correlates much more highly with UCOFT performance. At the very least we can conclude that prediction of gunnery performance is enhanced substantially by the addition of selected PAPB measures.

The present investigation also demonstrated that prediction of gunnery performance under degraded conditions does not appear to involve a different mix of abilities than is required for predicting performance when all fire control systems are operating. The same prediction model effectively predicted gunnery performance measures obtained under either mode.

Apart from its role in predicting UCOFT performance, in general the ASVAB CO composite demonstrated consistently high validities for predicting performance in a number of the OSUT performance domains. CO entered the regression model for the prediction of the TCGST, Military Stakes, Supervisor/Peer Ratings, and OSUT Performance Composites. In fact, except for the prediction of the Military Stakes composite, CO was the first and/or the only predictor to enter the regression equation for predicting each of the above composites.

In sum, the ASVAB together with the PAPB provide a very effective set of instruments for forecasting performance on a broad array of OSUT activities. Particularly useful are the ASVAB CO composite, the PAPB psychomotor Factor 1 scale, and the Combat scale from the PAPB ABLE. These three instruments should be given serious consideration as an additional source of information in aiding the process of selecting soldiers for participation in the Excellence Track.

REFERENCES

Administrator's Manual: Peer and Supervisor Rating Sessions for Concurrent Validation June - November 1985. (1985). Washington, D. C.: U.S. Army Research Institute for the Behavioral and Social Sciences.

Armored Vehicle Recognition (GTA 17-2-13). (October 1984). Headquarters, Department of the Army, Washington, D. C.

Arvey, R. D., Cole, D. A., Hazucha, J. F., & Hartano, F. M. (1985). Statistical power of training evaluation designs. Personnel Psychology, 38, 493-507.

Asher, J. J., & Sciarrino, J. A. (1981). A statistical power analysis for research designs in training. Unpublished manuscript, San Jose University.

Aviator's Recognition Manual (TC 1-88). (March 1977). U.S. Army Aviation Center, Fort Rucker, AL.

Bernstein, I. H. (1988). Applied multivariate analysis. New York: Springer-Verlag.

Bessemer, H. J. & Kraemer, R. E. (1979). The development and validation of audio-visual simulated performance tests using 35 mm slides (Research Memorandum 79-3). Fort Knox, KY; U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A077 976)

Biers, D. W., & Sauer, D. W. (1982). Job sample tests as predictors of M1 gunnery performance. Army Research Institute Research Report 584. (AD A139 433)

Black, B. A. (1980). ASVAB area score, Co, as a predictor of tank crew member performance (ARI Working Paper 80-9). Fort Knox, KY: U.S. Army Institute for the Behavioral and Social Sciences.

Black, B. A., & Mitchell, K. J. (1985). Predicting performance of M1 gunners (ARI Technical Report 707). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A176 649)

Campbell, C. H., & Black, B. A. (1982). Predicting trainability of M1 crewmen (Final Report MDA 903-80-C-0223). Fort Knox, KY: Human Resource Research Organization.

Campbell, (1986). Validation analysis for new predictors. U.S. Army Research Institute Research Report (Contract No. MDA903-82-C-053).

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation. Design and analysis issues for field settings. Chicago: Rand McNally College Publishing Company.

DeNisi, A. S., Cafferty, T. P., & Meglino, B. M. (1984). A cognitive view of the performance appraisal process: A model and research propositions. Organizational Behavior and Human Performance, 33, 360-396.

Eaton, N. K., Bessemer, D. W., & Kristiansen, D. M. (1979). Tank crew position assignment. Army Research Institute Research Report 391. (AD A077 841)

Eaton, N. K., Johnson, J., & Black, B. A. (1980). Job samples as tank gunnery performance predictors. Army Research Institute Research Report 473. (AD A100 973)

Eaton, N. K. & Whalen, D. E. (1980). Accuracy of tank gunnery scoring (ARI Working Paper 80-4). Fort Knox, KY: U.S. Army Research Institute for the Behavioral Sciences.

Graham, S. E. (1985). Assessing tank commander and gunner proficiency on U-COFT. Army Research Institute for the Behavioral and Social Sciences Conference Paper.

Graham, S. E., & Black, B. A. (1985). Defining and assessing tank commander excellence. Army Research Institute Report 1401. (AD A156 804)

Handbook for Sight Picture Training - M1 Tank (TC 17-12-1). (undated). Fort Knox, KY: US Army Research Institute for the Behavioral and Social Sciences.

Hough, L. M., Barge, B. N., Houston, J. S., McGue, M. K., Kamp, J. D., (1985, August). Problems, issues, and results in the development of temperament, biographical, and interest measures. Paper presented at the 93rd Annual Convention of the American Psychological Association. Los Angeles, California.

Kraemer, R. G. (1984). Fire Commands for the M1 Tank, Booklet 5: Single Target Engagements. RP 84-11-E. Fort Knox, KY: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A156 801)

Morrison, J. E., & Bessemer, D. W. (1981). Training and retention of armor machine gun tasks (Research Report 1317). Fort Knox, KY: U.S. Army Research Institute. (AD A128 824)

Morrison, J. E., & Bessemer, D. W. (1980). Observations and analysis of selected tasks in Armor One Station Unit Training. (ARI Working Paper 80-11). Fort Knox, KY: U.S. Army Research Institute for the Behavioral and Social Sciences.

Norusis, M. J. (1986). SPSS/PC+ for the IBM PC/XT/AT. Chicago: SPSSinc.

Phillips, R. (1985). CMF 19 active component personnel assessment. U.S. Army Armor Center, Fort Knox, KY.

Schmidt, F. L., Hunter, J. E., & Urry, V. W. (1976). Statistical power in criterion-related validation studies. Journal of Applied Psychology, 61, 473-485.

Seibert, J. H. (1987). The development of a knowledge test for armor trainees. Unpublished Master's Thesis, Western Kentucky University, Bowling Green, KY.

Siegel, A. I., & Bergman, B. A. (1975). A job learning approach to performance prediction. Personnel Psychology, 28, 325-339.

Steele, R. G. D., & Torrie, J. H. (1960). Principles and procedures of statistics. New York: McGraw Hill Book Co., Inc.

Toquam, J. L., Dunnette, M. D., Corpe, V., McHenry, J. J., Keyes, M. A., McGue, M. K., Houston, J. S., Russell, T. L., & Hanson, M. A. (1985, August). Development of cognitive/perceptual measures: Supplementing the ASVAB. Paper presented at the 93rd Annual Convention of the American Psychological Association. Los Angeles, California.

Unit-Conduct of Fire Trainer Instructor's Utilization Handbook (1985). General Electric Company: Daytona Beach, FL.

Visual Aircraft Recognition (FM 44-30). (October 1983). Headquarters, Department of the Army, Washington, D.C.

Wallace, J. R. 1982. The Gideon criterion: The effects of selection criteria on soldier capabilities and battle results. U.S. Army Recruiting Command (USARCPAE-RE).

Appendix

Intercorrelation Matrix: ASVAB, PAPB, UCOFT

Intercorrelations: ASVAB, PAPB, & Composite OSUT Criterion Performance Measures
(N = 123)

Correlations:	GT	GM	CL	MM	SC	CO	FA	OF	ST
GT	1.0000	.7733**	.9456**	.6132**	.8400**	.7089**	.8211**	.7007**	.8474**
GM	.7733**	1.0000	.8004**	.8604**	.8986**	.7862**	.7584**	.8411**	.9078**
CL	.9456**	.8004**	1.0000	.6003**	.7888**	.7000**	.8998**	.6647**	.8750**
MM	.6132**	.8604**	.6003**	1.0000	.8701**	.8614**	.6936**	.9534**	.7292**
SC	.8400**	.8986**	.7888**	.8701**	1.0000	.8795**	.7671**	.9032**	.8588**
CO	.7089**	.7862**	.7000**	.8614**	.8795**	1.0000	.8496**	.8636**	.7110**
FA	.8211**	.7584**	.8998**	.6936**	.7671**	.8496**	1.0000	.7196**	.8177**
OF	.7007**	.8411**	.6647**	.9534**	.9032**	.8636**	.7196**	1.0000	.7935**
ST	.8474**	.9078**	.8750**	.7292**	.8588**	.7110**	.8177**	.7935**	1.0000
FACTOR1	.3151**	.4034**	.3072**	.3501**	.3935**	.3496**	.3403**	.3843**	.4532**
FACTOR2	.4282**	.3258**	.3609**	.2357*	.3669**	.2662*	.2702*	.3049**	.3891**
FACTOR3	.0279	.1255	.0597	.1321	.1118	.1273	.0729	.1212	.0861
COMBAT	-.1829	-.1176	-.2136*	-.0745	-.0925	-.1087	-.2174*	-.0733	-.1363
DEPEND	.1699	.1387	.1881	.1123	.1241	.1293	.2169*	.1086	.1879
ACCURACY	.1440	.2826**	.1352	.2390*	.2485*	.2079	.1662	.1918	.2755*
LATENCY	.1324	.2700*	.1568	.2001	.2303*	.2372*	.1748	.1654	.2324*
SYSMGMT	.1995	.3342**	.2015	.2607*	.2771**	.2449*	.2136*	.2230*	.3188**
COFT	.1872	.3482**	.1939	.2748*	.2967**	.2706*	.2178*	.2280*	.3249**
TCGST	.2231*	.2932**	.2071	.2976**	.3117**	.3059**	.2285*	.3032**	.2683*
MILSTAKE	.5492**	.5741**	.4986**	.4885**	.5966**	.5096**	.4627**	.4986**	.5675**
RATING	.2801**	.2912**	.2976**	.2705*	.3005**	.3607**	.3318**	.2573*	.2491*
COMPOSIT	.4375**	.5353**	.4229**	.4728**	.5338**	.5128**	.4390**	.4562**	.5007**

Correlations:	FACTOR1	FACTOR2	FACTOR3	COMBAT	DEPEND	ACCURACY	LATENCY	SYSMGMT	COFT
GT	.3151**	.4282**	.0279	-.1829	.1699	.1440	.1324	.1995	.1872
GM	.4034**	.3258**	.1255	-.1176	.1387	.2826**	.2700*	.3342**	.3482**
CL	.3072**	.3609**	.0597	-.2136*	.1881	.1352	.1568	.2015	.1939
MM	.3501**	.2357*	.1321	-.0745	.1123	.2390*	.2001	.2607*	.2748*
SC	.3935**	.3669**	.1118	-.0925	.1241	.2485*	.2303*	.2771**	.2967**
CO	.3496**	.2662*	.1273	-.1087	.1293	.2079	.2372*	.2449*	.2706*
FA	.3403**	.2702*	.0729	-.2174*	.2169*	.1662	.1748	.2136*	.2178*
OF	.3843**	.3049**	.1212	-.0733	.1086	.1918	.1654	.2230*	.2280*
ST	.4532**	.3891**	.0861	-.1363	.1879	.2755*	.2324*	.3188**	.3249**
FACTOR1	1.0000	.6403**	.3002**	.0109	.1134	.2299*	.3462**	.4621**	.4083**
FACTOR2	.6403**	1.0000	.2354*	.0715	.0266	.1267	.3497**	.4581**	.3676**
FACTOR3	.3002**	.2354*	1.0000	-.1490	-.1573	.1521	.2156*	.2842**	.2563*
COMBAT	.0109	.0715	-.1490	1.0000	.0995	.1984	.2412*	.0877	.2051
DEPEND	.1134	.0266	-.1573	.0995	1.0000	-.0474	-.1404	-.0051	-.0743
ACCURACY	.2299*	.1267	.1521	.1984	-.0474	1.0000	.5725**	.5569**	.8340**
LATENCY	.3462**	.3497**	.2156*	.2412*	-.1404	.5725**	1.0000	.6252**	.8580**
SYSMGMT	.4621**	.4581**	.2842**	.0877	-.0051	.5569**	.6252**	1.0000	.8593**
COFT	.4083**	.3676**	.2563*	.2051	-.0743	.8340**	.8580**	.8593**	1.0000
TCGST	.1456	.1214	.0173	.0981	-.0617	.2774**	.2744*	.3507**	.3545**
MILSTAKE	.2737*	.3439**	.0414	-.0243	.0899	.3045**	.1645	.3567**	.3255**
RATING	.0261	.1380	-.0294	-.0456	.3335**	.0944	.1320	.2250*	.1779
COMPOSIT	.3084**	.3477**	.1058	.0875	.0973	.5490**	.5215**	.6483**	.6750**

Correlations:	TCGST	MILSTAKE	RATING	COMPOSIT
GT	.2231*	.5492**	.2801**	.4375**
GM	.2932**	.5741**	.2912**	.5353**
CL	.2071	.4986**	.2976**	.4229**
MM	.2976**	.4885**	.2705*	.4728**
SC	.3117**	.5966**	.3005**	.5338**
CO	.3059**	.5096**	.3607**	.5128**
FA	.2285*	.4627**	.3318**	.4390**
OF	.3032**	.4986**	.2573*	.4562**
ST	.2683*	.5675**	.2491*	.5007**
FACTOR1	.1456	.2737*	.0261	.3084**
FACTOR2	.1214	.3439**	.1380	.3477**
FACTOR3	.0173	.0414	-.0294	.1058
COMBAT	.0981	-.0243	-.0456	.0875
DEPEND	-.0617	.0899	.3335**	.0973
ACCURACY	.2774**	.3045**	.0944	.5490**
LATENCY	.2744*	.1645	.1320	.5215**
SYSMGMT	.3507**	.3567**	.2250*	.6483**
COFT	.3545**	.3255**	.1779	.6750**
TCGST	1.0000	.4036**	.3737**	.7620**
MILSTAKE	.4036**	1.0000	.2946**	.7150**
RATING	.3737**	.2946**	1.0000	.6510**
COMPOSIT	.7620**	.7150**	.6510**	1.0000

N of cases: 123 1-tailed Signif: *p< .01 **p< .001